ARMY RESEARCH LABORATORY

Performance Evaluation of a Semantic Perception Classifier

by Craig Lennon, Barry Bodt, Marshal Childers, Rick Camden, Arne Suppe, Luis Navarro-Serment, and Nicoleta Florea

ARL-TR-6653 **September 2013**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067

ARL-TR-6653 **September 2013**

Performance Evaluation of a Semantic Perception Classifier

Craig Lennon and Marshal Childers
Vehicle Technology Directorate, ARL

Barry Bodt
Computational and Information Sciences Directorate, ARL

Rick Camden and Nicoleta Florea
Engility Corporation

Luis Navarro-Serment and Arne Suppe
Carnegie Mellon University

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
September 2013	Final	1 July 2013–1 November 2013

4. TITLE AND SUBTITLE
Performance Evaluation of a Semantic Perception Classifier

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Craig Lennon, Barry Bodt, Marshal Childers, Rick Camden,* Arne Suppe, †
Luis Navarro-Serment, † and Nicoleta Florea*

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-VTA, RDRL-CII-C
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-6653

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
*Engility Corporation, 3750 Centerview Drive, Chantilly, VA 20151
†Carnegie Mellon University, The Robotics Institute/NSH-4621, 5000 Forbes Ave., Pittsburgh, PA 15213-3890

14. ABSTRACT
In order to assess the technology developed by members of the U.S. Army Research Laboratory's Robotics Collaborative Technology Alliance (RCTA), RCTA's Integration and Assessment team conducts experiments to test the limits of developed technologies. We conducted such an assessment in October 2013 to evaluate the performance of a semantic labeling system developed by researchers at Carnegie Mellon University. The system consists of a Basler AC1300gc camera with Fujinon TF28DA8 lens and processing software, which runs on a laptop computer. The apparatus functions on small robotic platforms, such as a K-bot. This report details the design for our data collection and our assessment of the technology by examining precision, recall, and *F* measure via exploratory data analysis and analysis of variance (ANOVA).

15. SUBJECT TERMS
semantic perception, robotics, Integrated Research Assessment, autonomous systems, RCTA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Barry Bodt
Unclassified	Unclassified	Unclassified	UU	78	19b. TELEPHONE NUMBER (*Include area code*) 410-278-6659

Contents

List of Figures

List of Tables

1. Introduction

One of the U.S. Army Research Laboratory's (ARL's) Robotics Collaborative Technology Alliance's (RCTA) research goals is to improve shared situational awareness between robots and humans. In order to develop an understanding of the environment and terrain common to humans and robots, one thread of the RCTA's research focuses on creating a system by which the robot can semantically classify objects in its environment.[1] Such a system, by enabling the robot to label objects in terms humans understand, would allow the robot to navigate under instructions given by a human in terms that another human would understand. A Soldier interacting with such a system might point to a building and say, "Go around to the back of that building and watch the rear door. Report if anyone comes out." The robot would then perform the following actions:

1. Face the direction in which the Soldier is pointing.
2. Identify the building in front of it as a building.
3. Recognize that the building has an opposite side out of view on which there is a door.
4. Develop a hypothesis about the position of that door and a plan for navigating to it.
5. Move to the intended position, adjusting its desired viewing position in light of things it sees while in route.
6. Navigate around or move any obstacles in route.
7. Identify a door found as the back door.
8. Determine that it has a good viewing position for seeing people come out the door.
9. Identify human beings exiting the building and report this to the Soldier.

In order to complete these steps, the robot must possess a means of identifying buildings, people, doors, and other objects as required by the mission and terrain.

An RCTA member, the Carnegie Mellon University (CMU) Robotics Institute, developed a system, based on recent classification methodology,[2] that takes a standard color digital photograph and classifies everything within it as one of the following: sky, tree, asphalt floor, grass, building, object, concrete floor, or gravel floor. In October 2012, this system was taken to Fort Indiantown Gap, PA, to be assessed by ARL's RCTA Integration and Assessment team. In

[1]Mitchell, R.; Bornstein, J. *Robotics Collaborative Technology Alliance (RCTA): FY12 Annual Program Plan*; U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, March 2012. Task P3 is static scene understanding.

[2]Munoz, D.; Bagnell, J. A.; Hebert, M. Stacked Hierarchical Labeling. *Proceedings of the 2010 European Conference on Computer Vision (ECCV)*, September 2010.

subsequent sections, we describe the data collected during the experiment and provide an assessment of the CMU system based on that data.

2. Data Collection

ARL's RCTA Integration and Assessment team conducted the experiment at the Combined Arms Collective Training Facility (CACTF) at Fort Indiantown Gap, PA, on 11–12 October 2012. It consisted of controlled data collection by CMU in the form of images of designated buildings taken from predetermined locations at different times of day. The images were then semantically labeled by the CMU classification system. We intended for the experiment to assess the system's ability to classify a building and its surroundings, and thus decided that every image would include a building. The experiment needed to be conducted without precipitation, which might damage the collection system. With these restrictions and based on input from the CMU system's developers, we determined that the major factors influencing the system's performance would be distance from the building, orientation with respect to the building, position of the sun, and the shape of the building and its background. Cloud cover might also influence the classification by changing the lighting conditions and creating shadows. Consequently, we included in our design the factors shown in table 1.

Table 1. Design factors for the data collection.

Factors	Levels
Distance from the building	5, 20, 35, and 50 m (when possible)
Orientation (angle to the building)	0°, ±30°, and ±60° (when possible)
Time of day	Morning, noon, and afternoon blocks
Building shape and background	Bar, church, cube, police station

The levels of building shape and background refer to the names of four buildings in the CACTF. Images of and data relevant to each building are contained in appendices A–D, with each building receiving its own appendix. The orientation refers to angles as perceived by a person whose back is placed at the point of the building toward which the robot would point, with forward as 0°, left as negative, and right as positive. Only for the church could the full range of angles and distances be collected. For other buildings, angles and distances were limited by the terrain. The distances and angles collected for each building, along with a satellite photo showing the positions, are included in each building appendix. The data collection was replicated twice according to the design described above, with the exception of the afternoon block, which was collected only once because of constraints on the time we could remain on site. The two collections, on 11 and 12 October, presented different levels of cloud cover—an uncontrolled factor that may influence the classification algorithm both via the presence of clouds and by its effect on lighting conditions.

The result of the data collection was 265 photos taken at 5 different time blocks from 53 locations. These photos were then labeled by CMU's classification system. In order to establish ground truth, we gave the same photos to a photographer who used *LabelMe*[3] software to classify everything in the image into the categories used by the CMU classification system: sky, tree, asphalt floor, grass, building, object, concrete floor, or gravel floor. The *LabelMe* software allows the user to divide an image into regions, using line segments to separate each region. Then the user gives the region a label. The result is that each pixel in the image is classified into one of the possible categories. Because of the requirement that regions be separated with straight line segments, there will be some pixels that are misclassified by the human user of *LabelMe*. These mislabeled pixels will be counted as incorrectly labeled by the CMU system, but we do not believe that this error has a large impact on the results, as the human mislabeling is likely small in proportion to the number of pixels in the images.

3. Data Analysis

3.1 Describing the Data and Performance Measures

Our collected data consisted of 265 images, with each pixel in the image given a label of sky, tree, asphalt floor, grass, building, concrete floor, or gravel floor by the classifier. Our ground truth consisted of the same 265 images, with each pixel given one of the same labels by a photographer. Thus, for each image, one can obtain the number of mislabeled pixels of each type and produce a confusion matrix. We aggregated these confusion matrices by design factors (or interactions of them), and based on this aggregate confusion matrix, we computed standard classification assessment statistics, such as precision, recall, and *F* measure.[4]

We explain our measures in the context of an example. Figures 1–3 show an image of the church, the classification provided via *LabelMe*, and the image shaded by the classifier, with regions colored according to the colors in table 2. The classification was mostly correct, with some trees classified as buildings and some buildings classified as sky. In table 2 and all subsequent tables, concrete, asphalt, and gravel designate concrete floor, asphalt floor, and gravel floor. Concrete in a wall is considered a building and labeled as such.

Comparing the *LabelMe* and classifier labeled images produces a confusion matrix, shown in table 3.

[3]Russell, B. C.; Torralba, A.; Murphy, K. P.; Freeman, W. T. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision* **May 2008**, *77* (1–3), 157–173.

[4]Fan, R. E.; Lin, C. J. A Study on Threshold Selection for Multi-Label Classification. http://www.csie.ntu.edu.tw/~cjlin /papers/threshold.pdf (accessed 14 May 2013).

Figure 1. An image of the church taken from location church 19.

Figure 2. A *LabelMe* classification of the church taken from location church 19.

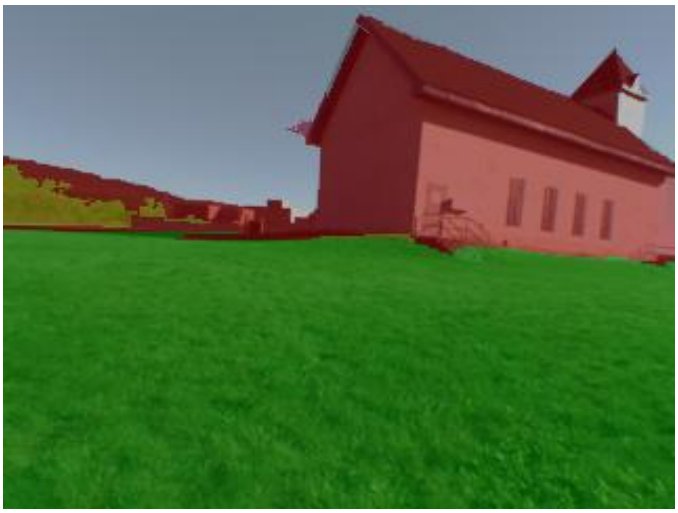

Figure 3. An image of the church, location church 19, classified by the CMU classifier.

Table 2. The labels of the colors used by the CMU classifier.

Sky		Tree		Asphalt		Grass	
Building		Object		Concrete		Gravel	

Table 3. Confusion matrix for the example church image (church 19).

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	17881	0	0	0	341	0	0	0
Tree	1	1096	0	17	787	0	0	0
Asphalt	0	0	0	0	0	0	0	0
Grass	0	38	0	40163	331	0	0	0
Building	565	51	0	114	14357	0	0	0
Object	0	14	0	12	368	0	0	0
Concrete	0	0	0	0	0	0	0	0
Gravel	0	0	0	0	0	0	0	0

Note: Entries represent counts of image pixels jointly classified through ground truth (rows) and semantic labeling (columns).

The labels for the columns indicate the pixel labels chosen by the CMU classifier, while the row labels indicate the photographer's labeling, which is our ground truth. For sky, 17881 pixels were labeled as sky by the classifier and the photographer, 565 pixels were labeled as sky by the classifier and as building by the photographer, and 341 pixels were labeled as building by the classifier and sky by the photographer. Anything off the diagonal of the matrix is considered an error of the classifier. Considering the first row, there were 18222 sky pixels, of which 17881 were correctly labeled by the classifier. This proportion (17881/18222) is the number of correctly classified sky pixels divided by the total number of sky pixels and is called the recall with respect to class sky.

Formally,

$$Recall_{class} = \frac{Number\ of\ class\ pixels\ correctly\ identified}{Total\ number\ of\ class\ pixels\ in\ the\ image}. \quad (1)$$

Summing down the column for sky, we see that the classifier identified 18447 pixels as sky, of which 17881 were so called correctly. This proportion is the precision for class sky:

$$Precision_{class} = \frac{Number\ of\ class\ pixels\ correctly\ identified}{Total\ number\ of\ pixels\ labeled\ as\ class}. \quad (2)$$

Thus $Recall_{sky} = 17881/18222$ and $Precision_{sky} = 17{,}881/18{,}447$. Recall can be increased at the expense of precision and vice versa, so the *F* measure takes the harmonic mean of the two as a compromise:

$$F_{class} = \frac{2}{\frac{1}{Precision_{class}} + \frac{1}{Recall_{class}}}. \quad (3)$$

$F_{sky} = 0.975$ for this image. We can evaluate the performance of the classifier across multiple classes by either a macro or micro average of the measures. The macro averaging is simply the sum of the *F* measure for each class divided by the total number of classes

$$F_{macro} = \frac{1}{Number\ of\ classes} \sum_{Classes} F_{class}, \quad (4)$$

and macro averaging for precision and recall is similar. Micro averaging requires us to compute recall and precision for the whole confusion matrix:

$$Precision_{micro} = \frac{\sum_{Classes} Number\ of\ class\ pixels\ correctly\ identified}{\sum_{Classes} Total\ number\ of\ pixels\ labeled\ as\ class}. \quad (5)$$

$$Recall_{micro} = \frac{\sum_{Classes} Number\ of\ class\ pixels\ correctly\ identified}{\sum_{Classes} Total\ number\ of\ class\ pixels\ in\ the\ image}. \quad (6)$$

Since all pixels in our image must be labeled, micro precision and micro recall are the sum of the diagonal elements of the confusion matrix divided by the total number of pixels and will be the same. The micro *F* measure is equal to precision and to recall in our case, but it is generally defined as the harmonic mean of micro precision and recall

$$F_{micro} = \frac{2}{\frac{1}{Precision_{micro}} + \frac{1}{Recall_{micro}}}. \quad (7)$$

For our example from the church, the results are found in table 4.

Table 4. Performance measures for church example photo.

Precision Macro	Recall Macro	*F* Macro	Precision Micro	Recall Micro	*F* Micro
0.941	0.875	0.898	0.965	0.965	0.965

The macro *F* measure weights each class equally, while the micro *F* measure weights the value of each class by the number of pixels it contributes to the total. Thus weak performance on rarely appearing classes will be noticeable in low macro *F* values combined with high micro *F* values.

In addition to computing these measures for an image, we can aggregate the confusion matrices either by adding together all the confusion matrices for a measure of overall performance or by adding together only those for images taken under specific conditions, such as at a location, for a part of the day, for a building, or under conditions with glare from the sun. We will do this in the next section, beginning with the highest levels of aggregation and proceeding to lower levels to evaluate the CMU classifier under different conditions.

3.2 Evaluation

We begin with the aggregate confusion matrix and performance measures for all 265 images. The confusion matrix and performance measures are provided in tables 5 and 6.

Table 5. Performance measures for all 265 images.

Macro Precision	Macro Recall	Macro *F*	Micro *F*
0.806	0.776	0.776	0.916

Table 6. Confusion matrix for all 265 images.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	5991945	4973	8	294	201973	1063	91	0
Tree	46973	265023	190	27349	73045	1393	465	104
Asphalt	1626	33	1851695	314	37036	251	64920	7959
Grass	4202	17062	2683	4359858	265808	17339	67993	101011
Building	169012	6326	314	29112	4415157	8279	28382	9
Object	8278	3870	617	49450	226896	9552	40719	45
Concrete	804	219	91070	3841	36799	221	1131451	544
Gravel	912	472	6181	16103	10266	1253	5124	324983

Note: Entries represent counts of image pixels jointly classified through ground truth (rows) and semantic labeling (columns).

A glance at the confusion matrix shows that while the classifier is usually right for most classes, it is usually wrong with regard to objects. It labels objects as buildings more often than as objects, and pixels with object labels are more likely to be grass than objects. Since the low overall macro *F* statistic may derive from poor performance in one class, we include the *F* measures for each class in table 7.

Table 7. Class *F* measures for all 265 images.

Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
0.965	0.744	0.946	0.935	0.890	0.050	0.869	0.813

Table 7 confirms our suspicion that the classifier is much worse at the objects than at any other class. This is not surprising, because the object class includes everything not identified as one of the other classes, from generators to sandbags. The object is like a label of "other," and a class composed of objects with little in common might be hard to identify. With the object class removed, the overall macro *F* measure would be 0.88.

Having examined the overall performance of the classifier, we now compare it across conditions of different sunlight, and then across the design factors presented in table 1. We only present the statistical performance measures for these conditions, relegating the confusion matrices to appendix E. In some images, such as figure 4, the sun produced a glare that may have interfered with the image. The classification of images into sun and no-sun categories was done by the CMU experimenters; 54 images were classified as sun. In table 8 we see that all performance measures are worse in the presence of sun interference.

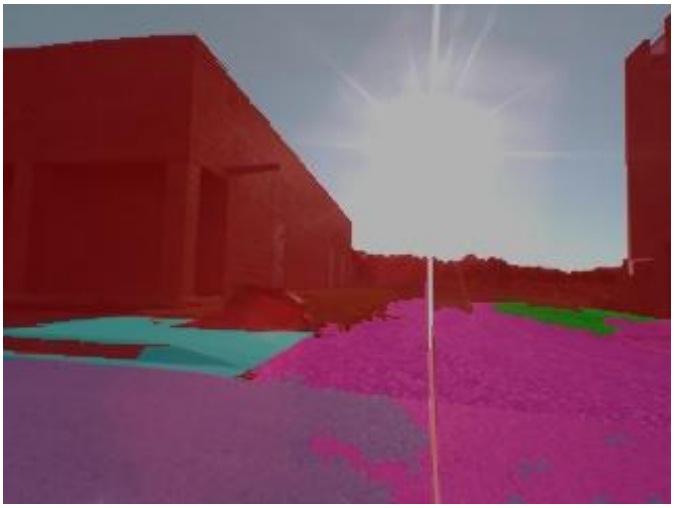

Figure 4. An example of sun interference.

Table 8. Performance measures for sun and no-sun images.

	Macro Precision	Macro Recall	Macro *F*	Micro *F*
No-Sun	0.828	0.782	0.778	0.926
Sun	0.787	0.736	0.749	0.87

Table 9 shows a breakdown of how the sun affected the *F* measure for each class, and it is interesting to see that there is not uniform degradation across classes. Some classes are worse with sun interference, and others are not. Of course, not all locations were equally likely to have sun interference, and in those locations where it cannot occur, sun interference will have no influence. It may be the case that the performance in the building class suffers so much under the sun because, as vertical surfaces, buildings have more problems with over / under exposure.

Table 9. *F* measure by class for all images by sun interference.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
No-Sun	0.963	0.746	0.957	0.953	0.911	0.045	0.848	0.800
Sun	0.970	0.731	0.828	0.839	0.793	0.068	0.924	0.843

Next, we consider the influence of the design factors beginning with the building. Table 10 contains the performance measures for each building, and table 11 breaks down the *F* measures by class for each building. There is no consistent trend across classes by building, though the bar and cube have classes in which they performed much worse than others. The entries labeled "None" indicate the absence of any pixels of that class in the ground truth image.

Table 10. Performance measures by building.

	Macro Precision	**Macro Recall**	**Macro *F***	**Micro *F***
Cube	0.801	0.750	0.752	0.947
Church	0.764	0.771	0.761	0.907
Bar	0.770	0.716	0.710	0.909
Police	0.817	0.714	0.744	0.903

Table 11. *F* measures by class by building.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Cube	0.980	0.788	0.962	0.982	0.896	0.049	0.842	0.517
Church	0.959	0.776	0.979	0.926	0.858	0.010	0.818	None
Bar	0.964	0.517	0.912	0.921	0.918	0.055	0.475	0.915
Police	0.954	0.718	0.928	0.717	0.896	0.084	0.913	None

We have also aggregated confusion matrices by the time of day during which the image was taken, with the morning session starting around 0930, the midday session around 1300, and the afternoon session around 1630. The performance measures for this aggregation are in table 12, with table 13 showing the breakdown by class and time of day. Again, there are no striking patterns in these tables.

Table 12. Performance measures by day and time of day.

	Macro Precision	Macro Recall	Macro *F*	Micro *F*
Morning 11th	0.797	0.795	0.785	0.906
Midday 11th	0.816	0.795	0.784	0.925
Afternoon 11th	0.812	0.747	0.756	0.903
Morning 12th	0.813	0.784	0.79	0.926
Midday 12th	0.811	0.769	0.768	0.918

Table 13. *F* measures by class by time.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Morning 11th	0.941	0.746	0.963	0.937	0.845	0.072	0.923	0.854
Midday 11th	0.981	0.825	0.963	0.934	0.902	0.038	0.880	0.745
Afternoon 11th	0.952	0.665	0.906	0.946	0.881	0.023	0.873	0.805
Morning 12th	0.985	0.792	0.929	0.925	0.912	0.102	0.828	0.850
Midday 12th	0.960	0.709	0.966	0.931	0.905	0.029	0.832	0.813

Our last aggregation is by distance and orientation. Since different buildings have different distances and orientations, we present here the interaction between building, distance, and orientation. We examine only the macro *F* measures for each interaction in this section. The appendices for each building contain plots for the micro measures. In the plots shown in figures 5–8, each number represents the macro *F* measure of the five runs done at the indicated location. The vertical and horizontal axes indicate meters parallel and perpendicular to the focal point on the wall of the building. On each plot are two labeled locations that can be compared with figures A-2, B-2, C-2, and D-2 in the building appendices. The lowest 10th percentile of macro *F* measures is at 0.636, so we have colored results below this number red to distinguish them. For micro *F* measure, the lowest 10th percentile was 0.846, and for the micro *F* plots in the appendices, we have colored numbers below this red. There does not seem to be any obvious relation between distance and orientation, except perhaps as it relates to sun, since sun interference can only occur from some distances and orientations.

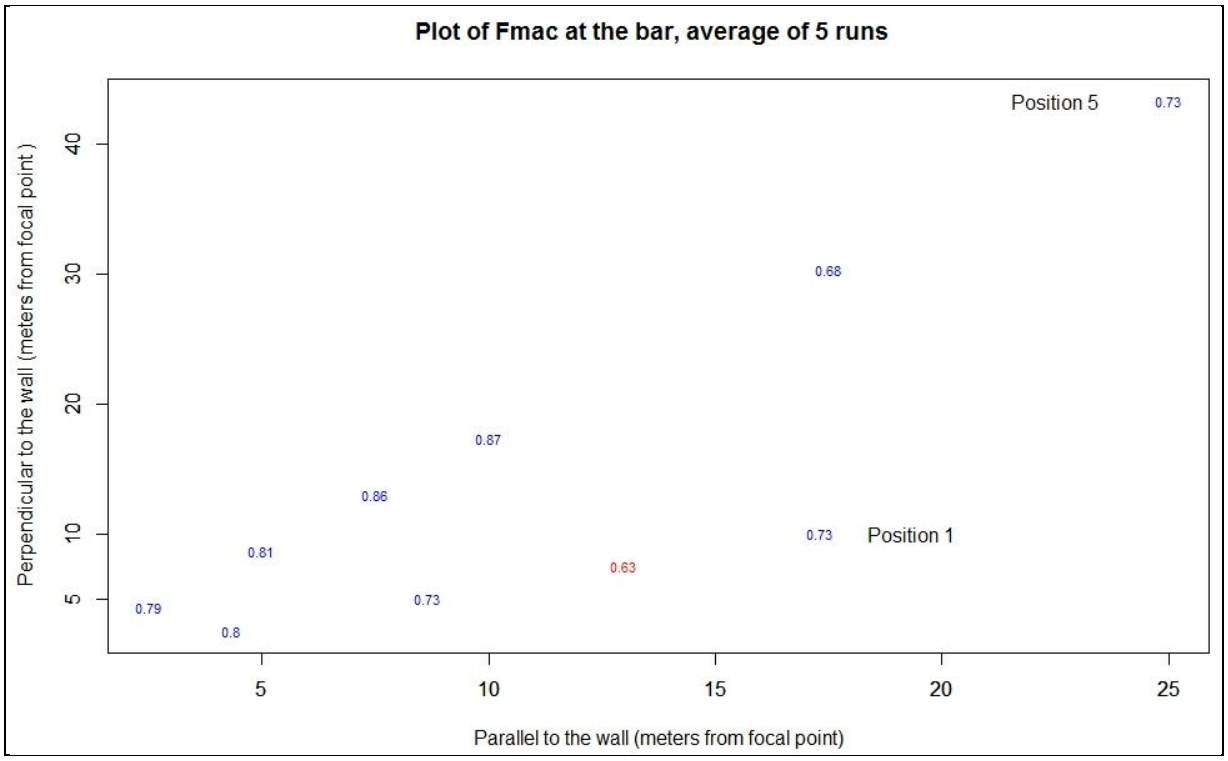

Figure 5. Plot of macro *F* measure for each location at the bar.

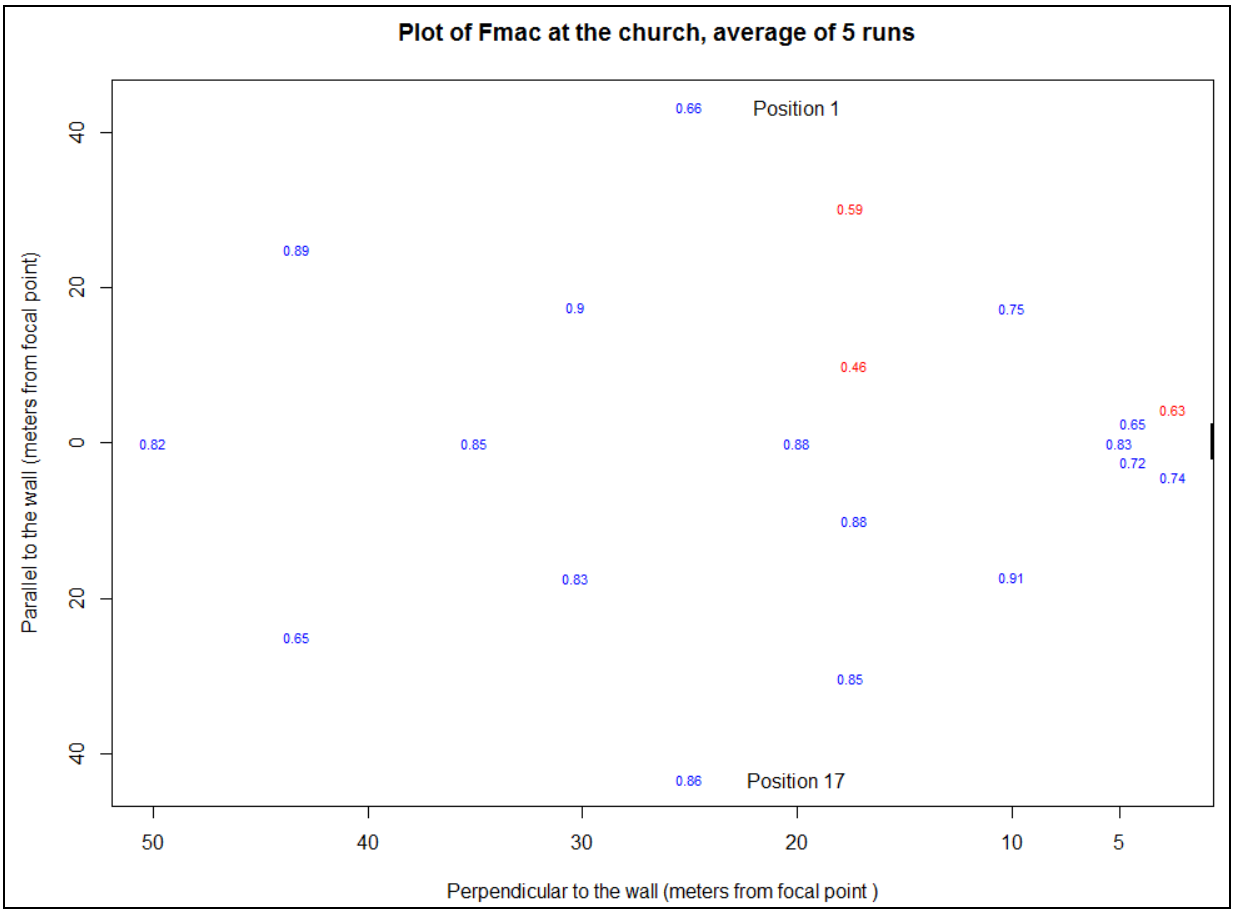

Figure 6. Plot of macro *F* measure for each location at the church.

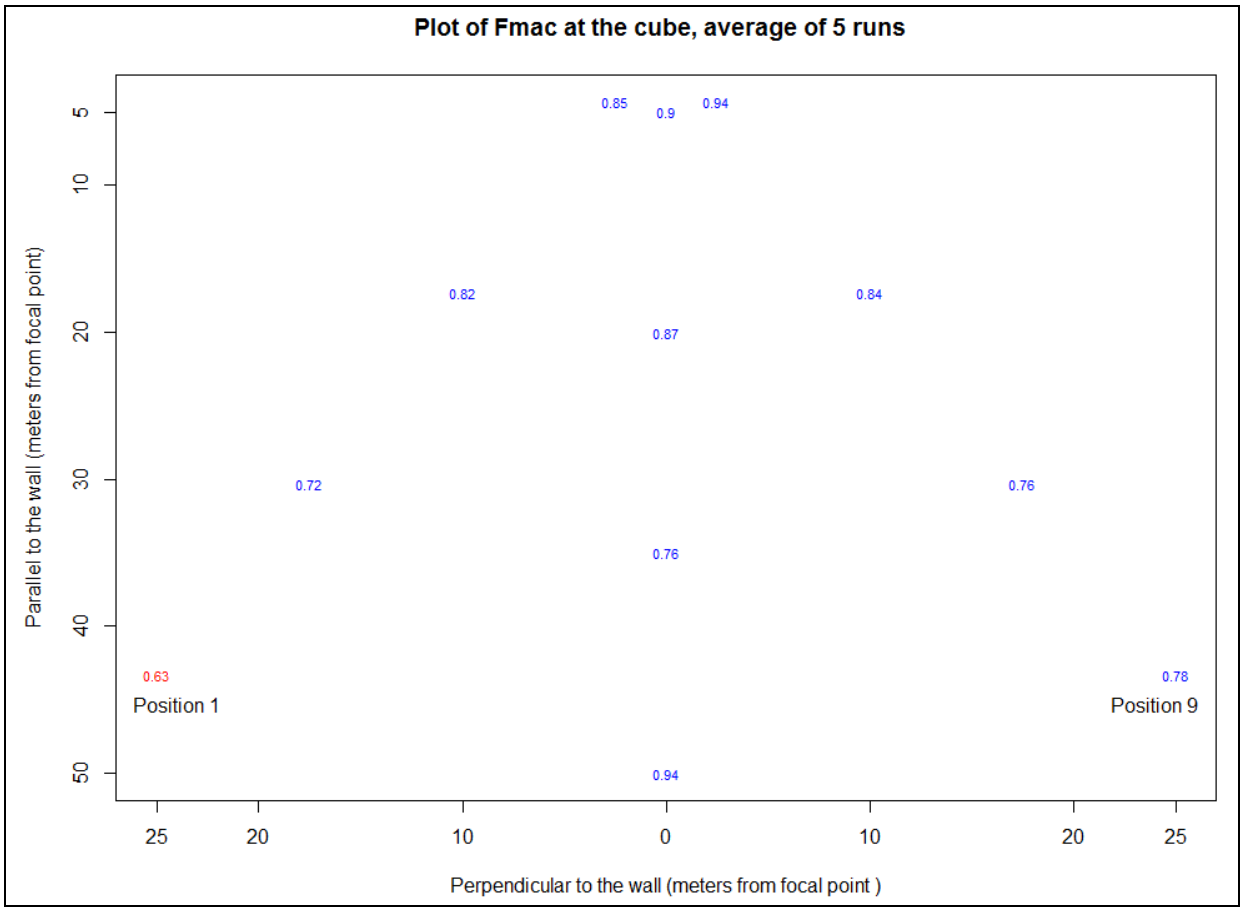

Figure 7. Plot of macro *F* measure for each location at the cube.

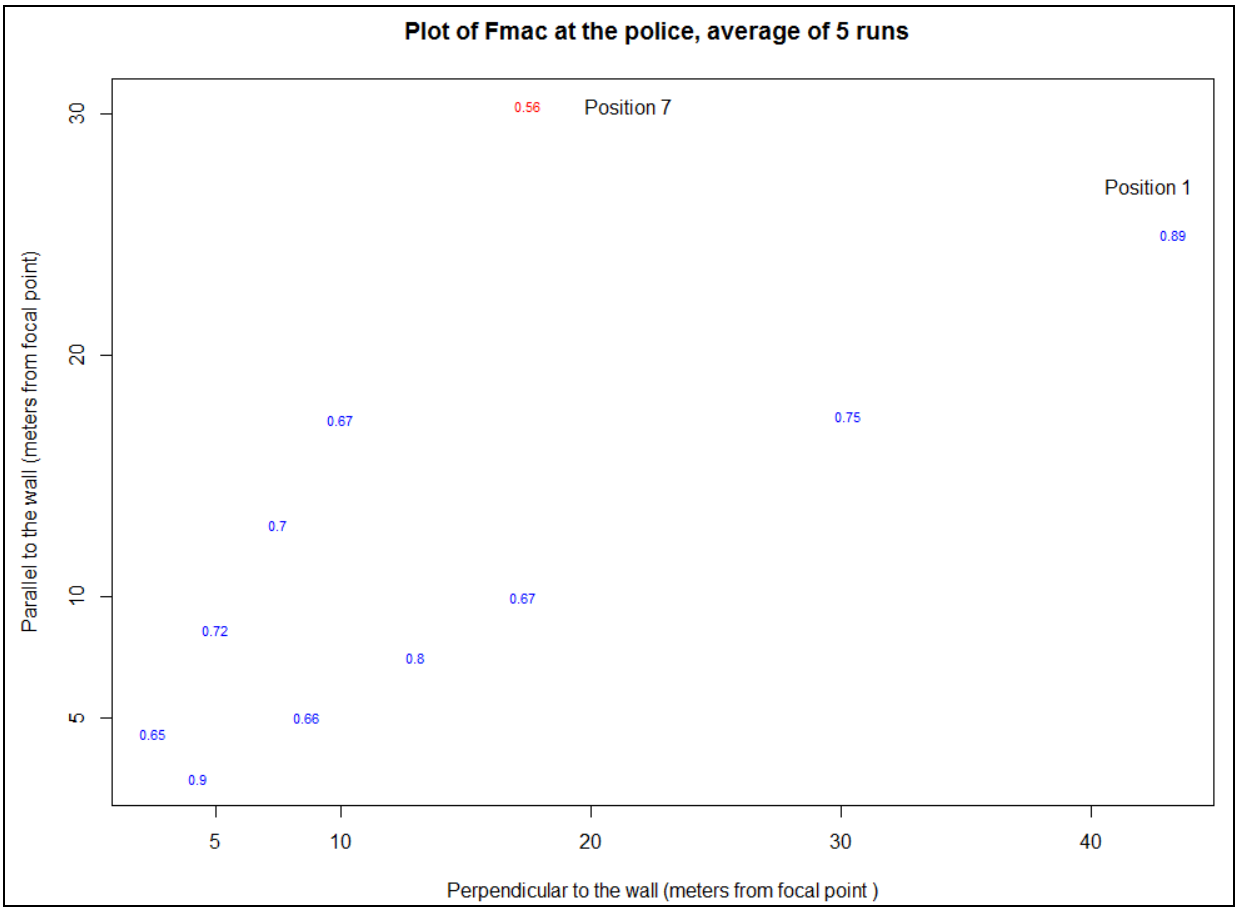

Figure 8. Plot of macro *F* measure for each location at the police station.

Having searched for trends in the macro and micro *F* measure by design factor, we are left to note that no obvious patterns appear. This does not, however, preclude the possibility of patterns for individual classes. It would take too long to conduct a detailed examination of every class, and we would likely find some spurious patterns in the process. Given the mission scenario mentioned in the introduction, it may be important to recognize a building when one is present in the image. Consequently, we devote the next section to an examination of how the classifier performed with respect to the building class.

3.3 Performance on Buildings

Identifying a building was one of the central requirements of the mission scenario under which we were testing. Three of the design factors in table 1 are descriptions related to a building (distance, angle, type). Thus, it may be that classifier performance is best represented by how well the classifier deals with the building class. Low precision with respect to the building class means perceiving building pixels where none are present, and low recall means failing to identify building pixels when they do exist. Rather than average these measures via the *F* measure, we examine them separately. We begin with plots of the cumulative distribution of building precision and recall for all images, which are in figures 9 and 10. Comparing the two shows

recall to be substantially better than precision, suggesting that the classifier, when in doubt, errs on the side of saying a building is present.

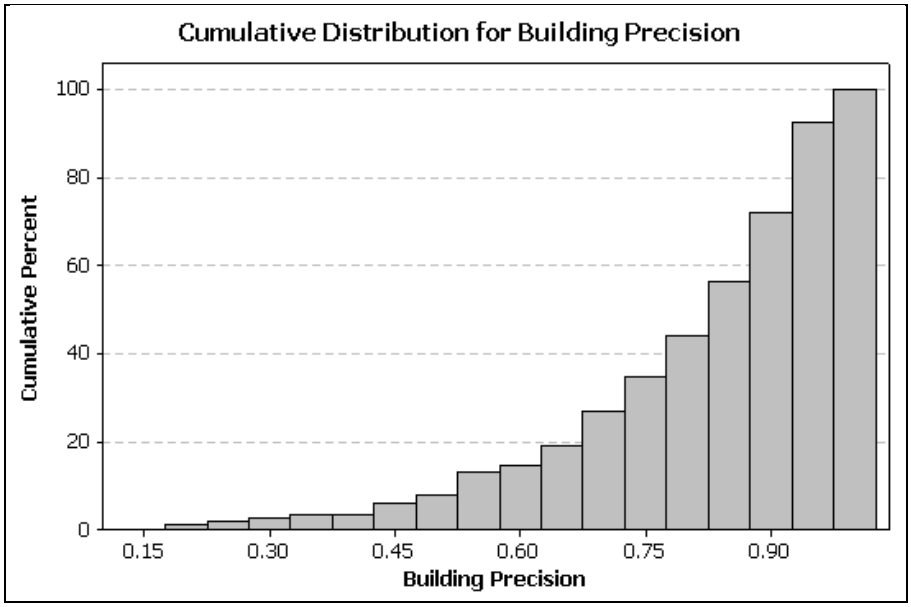

Figure 9. Precision for the building class (for all images).

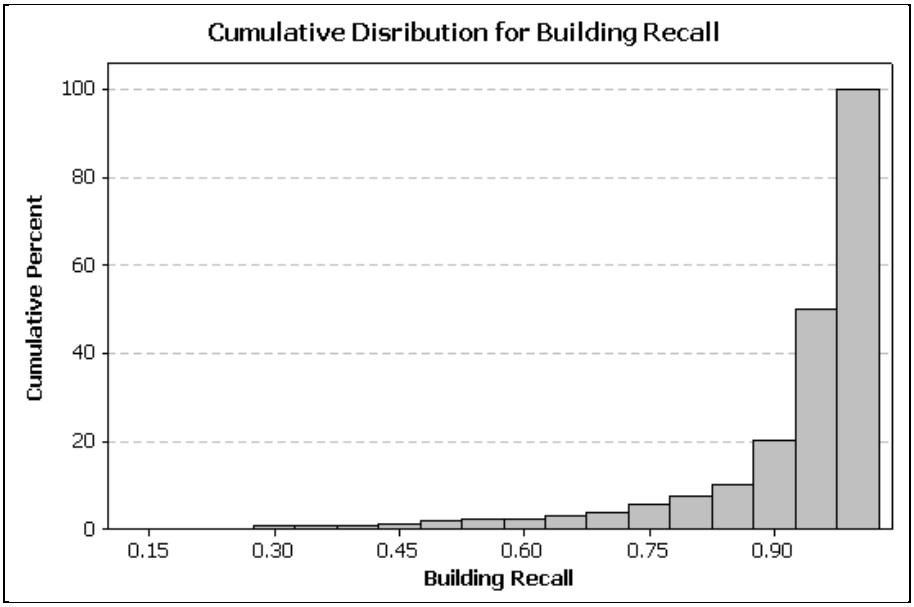

Figure 10. Recall for the building class (for all images).

Since the building class precision and recall may well depend on which building we consider, we look at the effects of distance and angle by building. Figures 11 and 12 show building precision by location and distance and by location and angle, while figures 13 and 14 show building recall broken down in the same way.

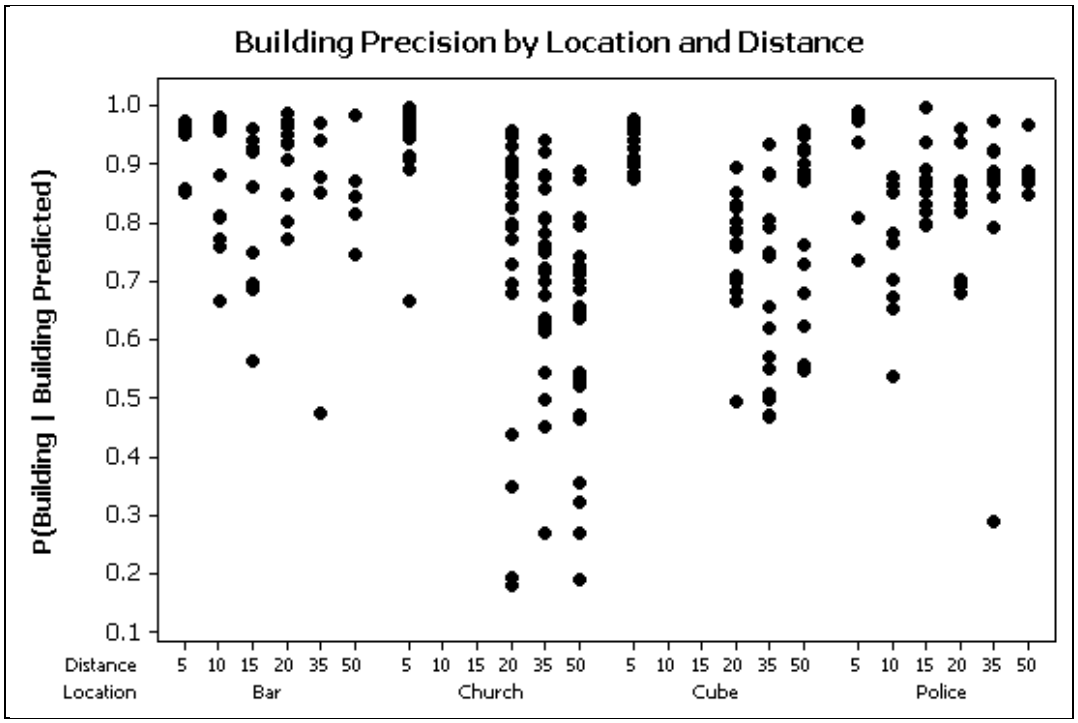

Figure 11. Building precision by location and distance.

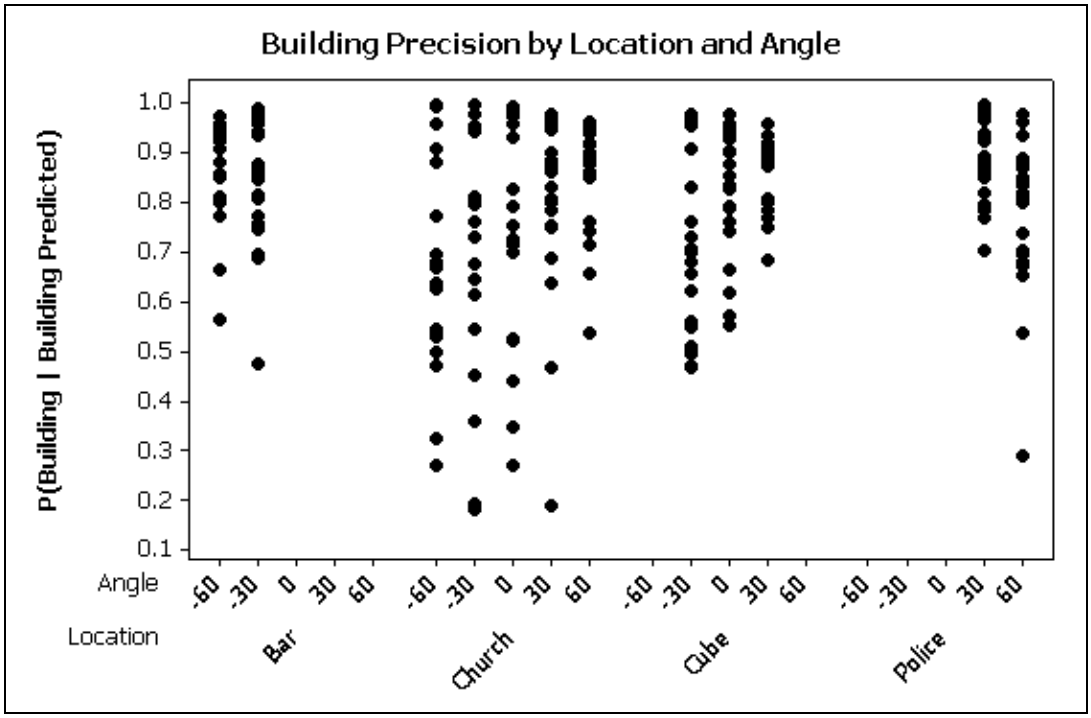

Figure 12. Building precision by location and angle.

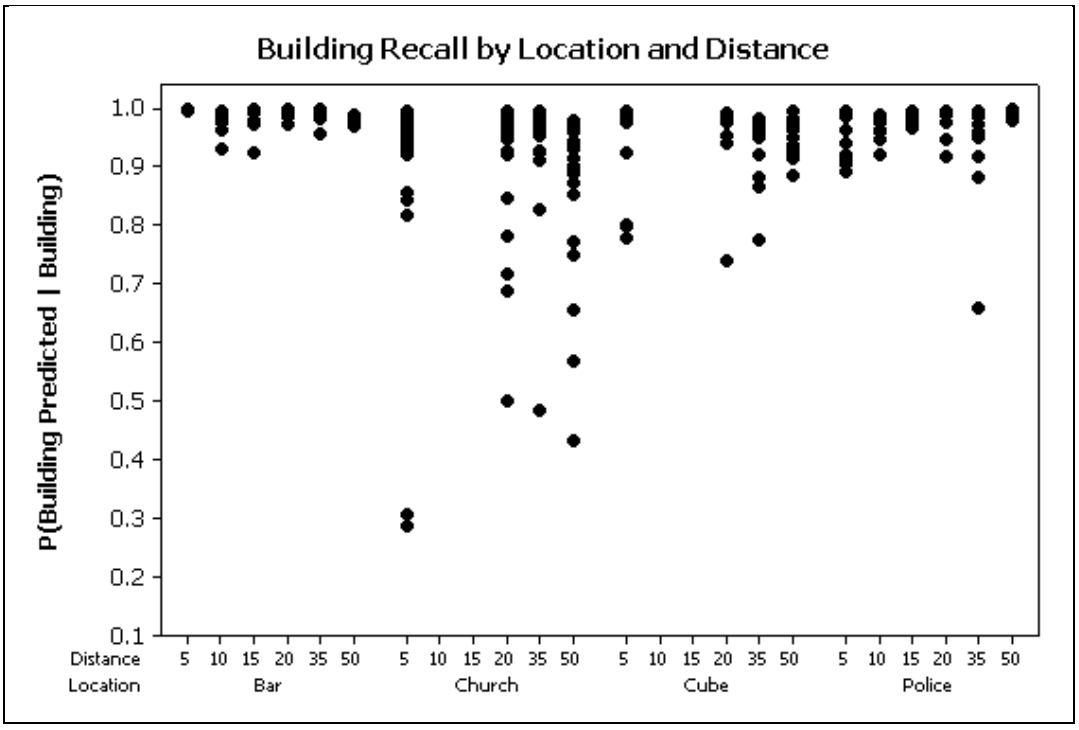

Figure 13. Building recall by location and distance.

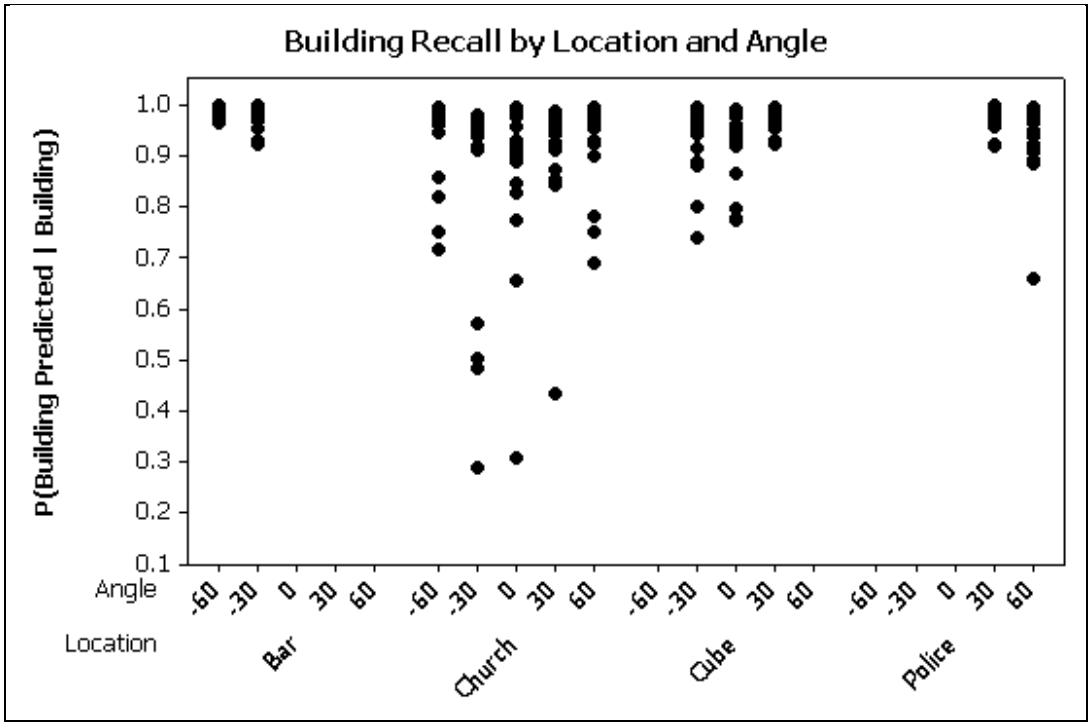

Figure 14. Building recall by location and angle.

Based on these plots, building recognition seems to perform worst at the church. Analysis of variance (ANOVA) results given in appendices A–D suggest that distance and angle affect precision at all building locations except at the bar, where angle has no effect. Sun effects are potentially captured by time of day, which was shown by the ANOVA to be statistically significant for all building locations except at the cube where the northerly camera view kept the sun behind or to the side during data collection for all time periods. The impact of design factors on recall was inconsistent, and the effect sizes were small. Because distance from a building could be confused with the effects of sun interference, we examine the effects of distance at the church with and without sun interference.

For example, in figure 15, we see the mean precision broken down by sun interference and no-sun interference. The bars extend from the circles to one standard error. Few images had sun interference, so the absence of error bars for sun interference measurements probably indicates only one sun interference image taken at the site, while the absence of error bars for no-sun images indicates a tight grouping of several measurements around the mean. The negative influence of sun artifacts makes the time-of-day effect statistically significant, as seen in appendix B, and distance and angle also affect the response.

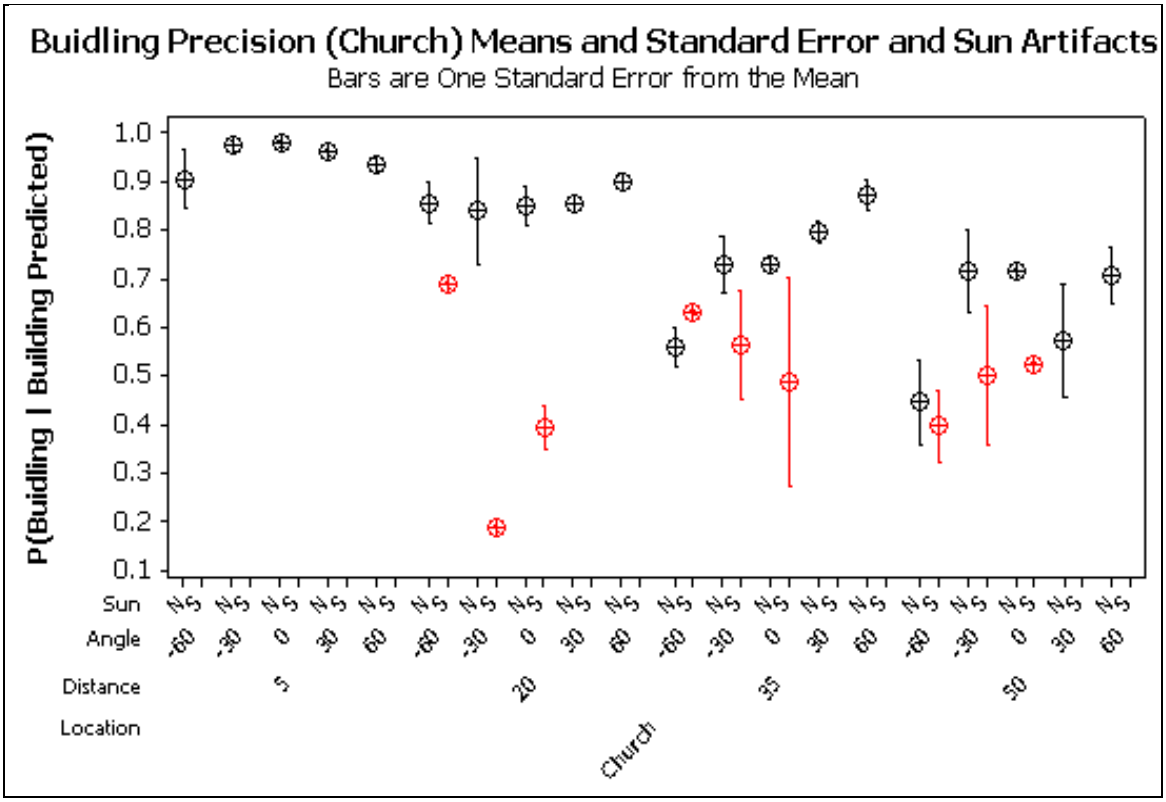

Figure 15. Plot of building precision at the church.

Note: The circles indicate means, with bars indicating one standard error. Red circles indicate images with sun interference, and black represents means of images without sun interference.

We are not certain why the classification of the sun images is worse than that of the no-sun images, but one conjecture is that it relates to over- and under-exposure in the images. For example, many images of the church were taken with one side in shadow and one side lit by the sun. Perhaps this causes the church images to be over-exposed, and this, in turn, degrades the performance of the classifier.

Building precision might decrease with distance because of the background. As you move farther back from the church, there are no other buildings in the background, whereas if you move farther back from the police station or bar, there are other buildings in the background. So for the church and cube, moving away from the building means fewer building pixels in the image. With fewer building pixels, mislabeling has a greater effect on precision and recall. To see the number of building pixels present, compare the images of the church, police station, and cube at 50 and 20 m, shown in figures 16–18. There are many more building pixels in the police station at 50 m than in the church at 50 m or cube at 20 m.

Figure 16. Church at 50 m.

Figure 17. Cube at 20 m.

Figure 18. Police station at 50 m.

It may be that the drop in precision we see with distance is a drop corresponding to fewer building pixels in the image.

Plots of building precision and recall by distance and location, with and without sun interference for each building, are included in the building appendices. They can be summarized as showing that building precision is worse with sun for all buildings, but that building recall is *not* worse with sun interference. It seems that sun contributes more to false positives than to false negatives. In the next section, we take a more detailed look at the images on which the classifier performed poorly. We return to examining macro and micro *F* measures, but this time we examine the individual images on which performance on these images was worst.

3.4 Examination of Selected Images

We can see how the micro and macro *F* measures of individual images are distributed by considering a plot of the empirical distribution of micro and macro *F* measure. In figures 19 and 20, the vertical axis is the proportion of the data less than the value shown along the horizontal axis. Most images have values of greater than 0.7 and 0.8 for macro and micro *F* measures, respectively. To gain insight into why some images have extremely low *F* measures, we selected the 10 worst images with respect to micro and macro *F* measure.

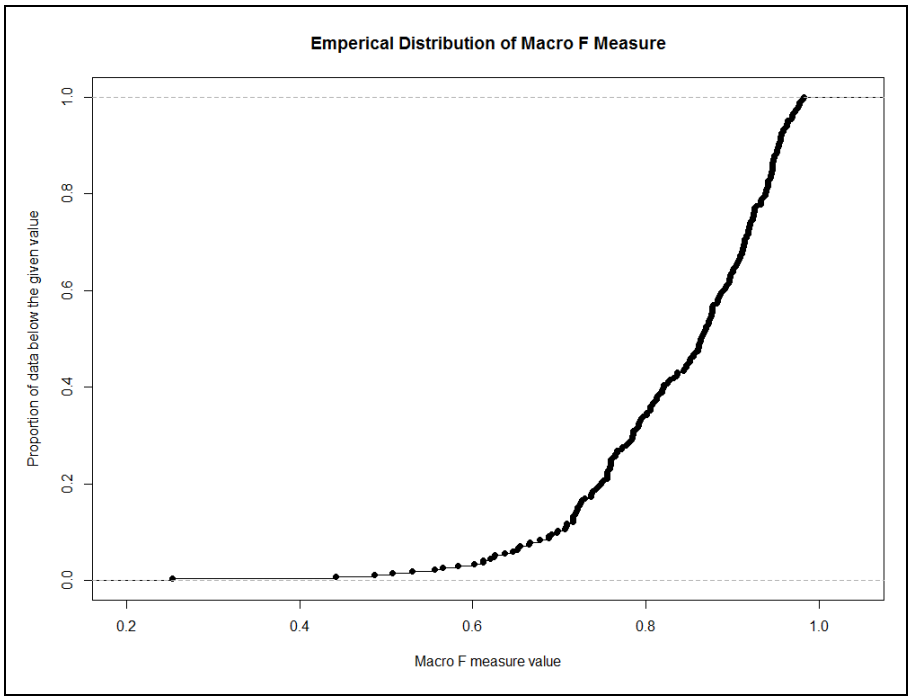

Figure 19. The empirical distribution for macro *F*.

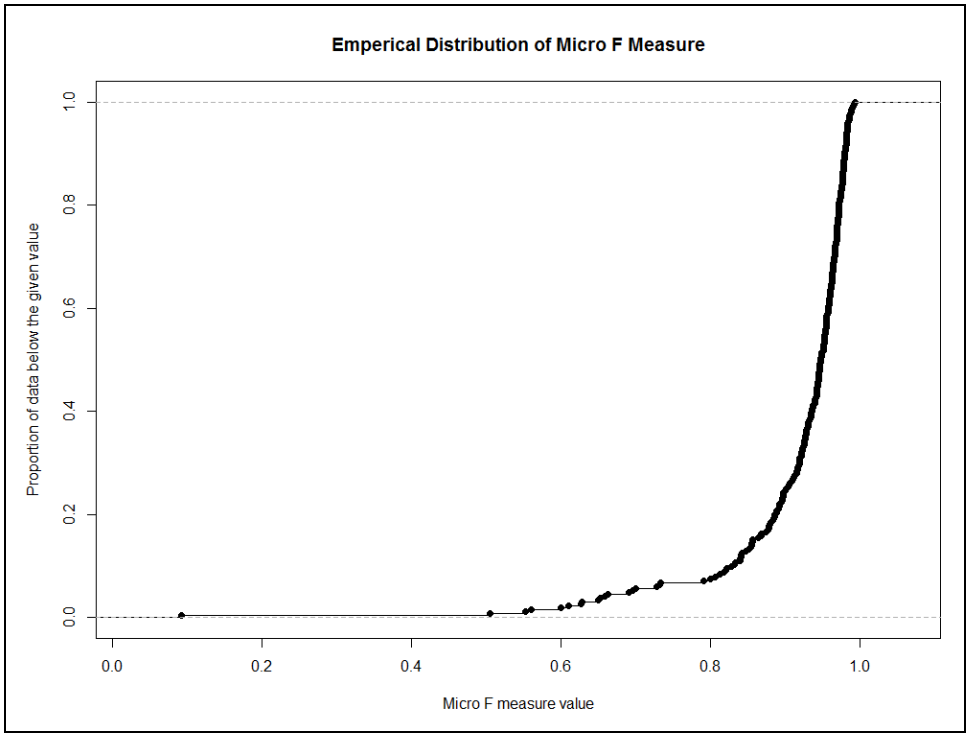

Figure 20. The empirical distribution for micro *F*.

Since many images were in the 10 worst for both micro and macro *F* measure, we are left with 12 images. We present a selection of these images in figures 21–25, noting the major source of error in each image. The remaining images are in appendix F.

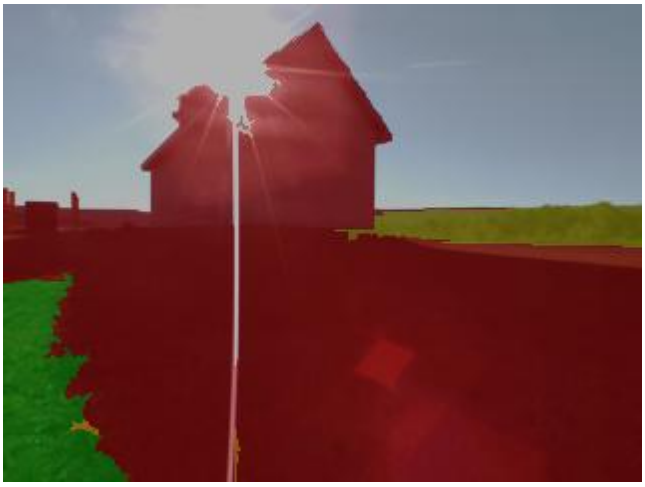

Figure 21. Church 7 on the morning of 11 October. The major error is that grass is classified as building.

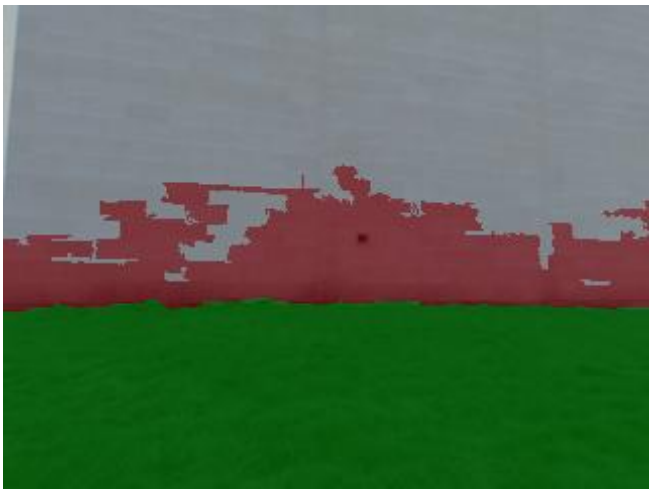

Figure 22. Church 8 on the morning of 11 October. Here much of the building is classified as sky.

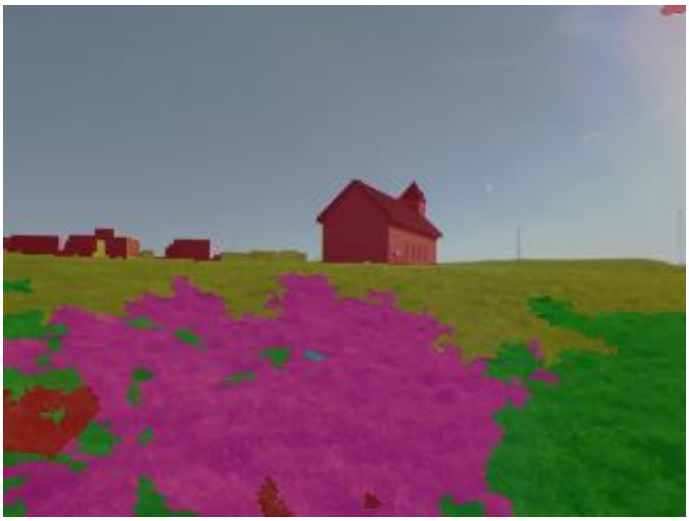

Figure 23. Church 13 on the morning of 11 October. Grass is classified as tree, asphalt, and building.

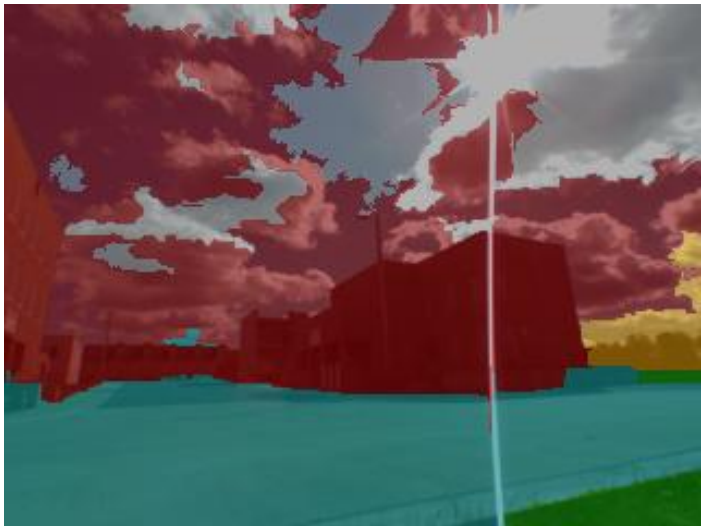

Figure 24. Police 7 around noon of 12 October. Sky is classified as building and objects.

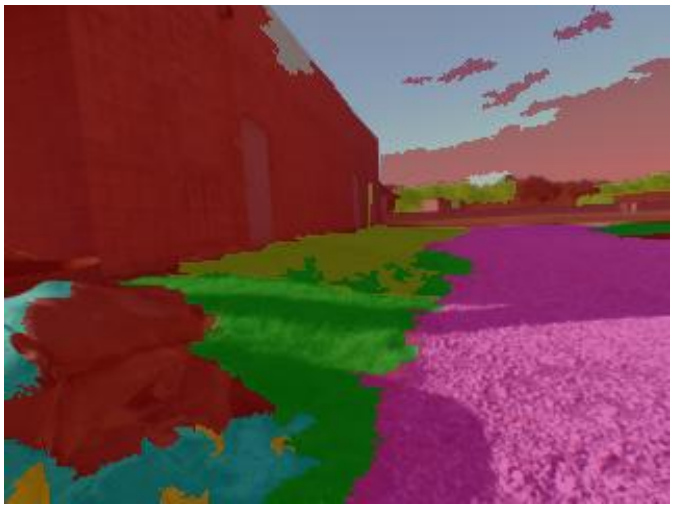

Figure 25. Bar 2 on the morning of 11 October. Grass is classified as trees, sky and tree are classified as building, and an object as building and concrete floor.

While there is no obvious pattern to the cause of the errors in the worst images, the result is large contiguous segments of the image being mislabeled. When the classifier makes errors, at least in these worst images, they seem to be concentrated in a few regions and not scattered uniformly about the image. This knowledge may be useful for the purpose of mitigating the effects of errors in the classifier.

4. Conclusions

The CMU classifier accurately segments and labels most pixels in most images. When failing, it tends to classify a contiguous, sizeable part of an image as being of the wrong class. Consequently, the classifier seems to either perform well or make large concentrated mistakes. With respect to the classification of buildings, we believe precision in the building class is lower when there are fewer building pixels in an image, meaning that recognition of a specific building will be worse when one is farther from that building. We also believe that building precision is decreased by sun interference, but that building recall is not appreciably affected. Performance in the building class might be improved with a method for recognizing and eliminating sun interference. Sun interference is not a major contributor in across-class measures of performance (macro/micro *F* measures). In these cross-class measures, the classifier does not consistently fail under a given set of circumstances but fails infrequently under disparate conditions. This failure pattern might be one that can be mitigated in practical use by averaging the results of images taken over time, so that objects receive their most frequent label. With such a scheme, the cost of infrequent large failures may be reduced.

Until precision is improved, the intelligence structure of the robot must make allowances for building false positives. Perhaps buildings recorded by the semantic classifier could be coded as hypothetical buildings to be confirmed by repeated viewings. The good performance of recall means the semantic classifier is less likely to “eliminate” buildings that are present, at least in its current configuration.

INTENTIONALLY LEFT BLANK.

Appendix A. The Bar

Figures A-1 through A-9 and tables A-1 through A-3 show details and results for the bar building.

Figure A-1. The bar is the red building on the left.

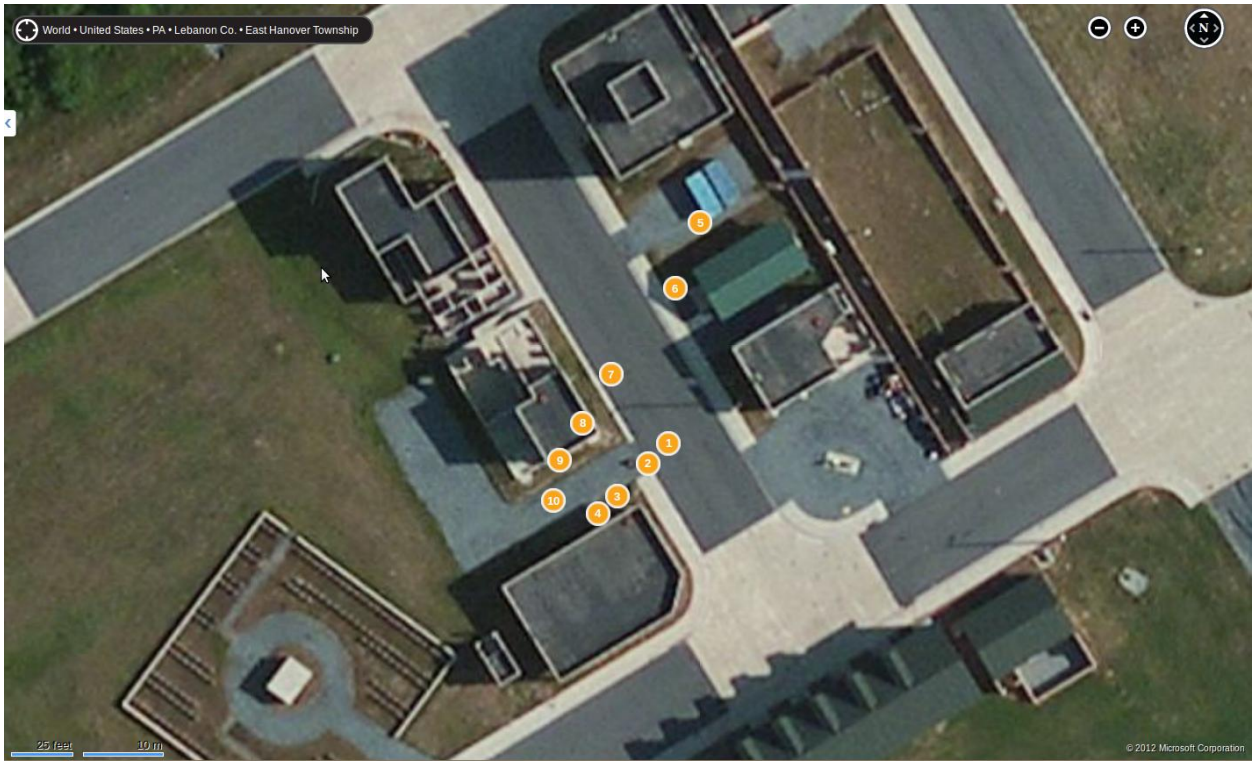

Figure A-2. Data collection positions around the bar.

Table A-1. *F* measure by distance and angle around the bar.

Position	Angle (°)	Dist (m)	Fmac	Fmic	P Bldg	R Bldg
1	60	20	0.732	0.918	0.873	0.992
2	60	15	0.634	0.743	0.784	0.987
3	60	10	0.734	0.910	0.857	0.983
4	60	5	0.795	0.945	0.917	0.995
5	30	50	0.731	0.950	0.852	0.979
6	30	35	0.677	0.894	0.822	0.984
7	30	20	0.866	0.950	0.941	0.991
8	30	15	0.856	0.913	0.791	0.973
9	30	10	0.815	0.925	0.857	0.973
10	30	5	0.788	0.941	0.944	0.996

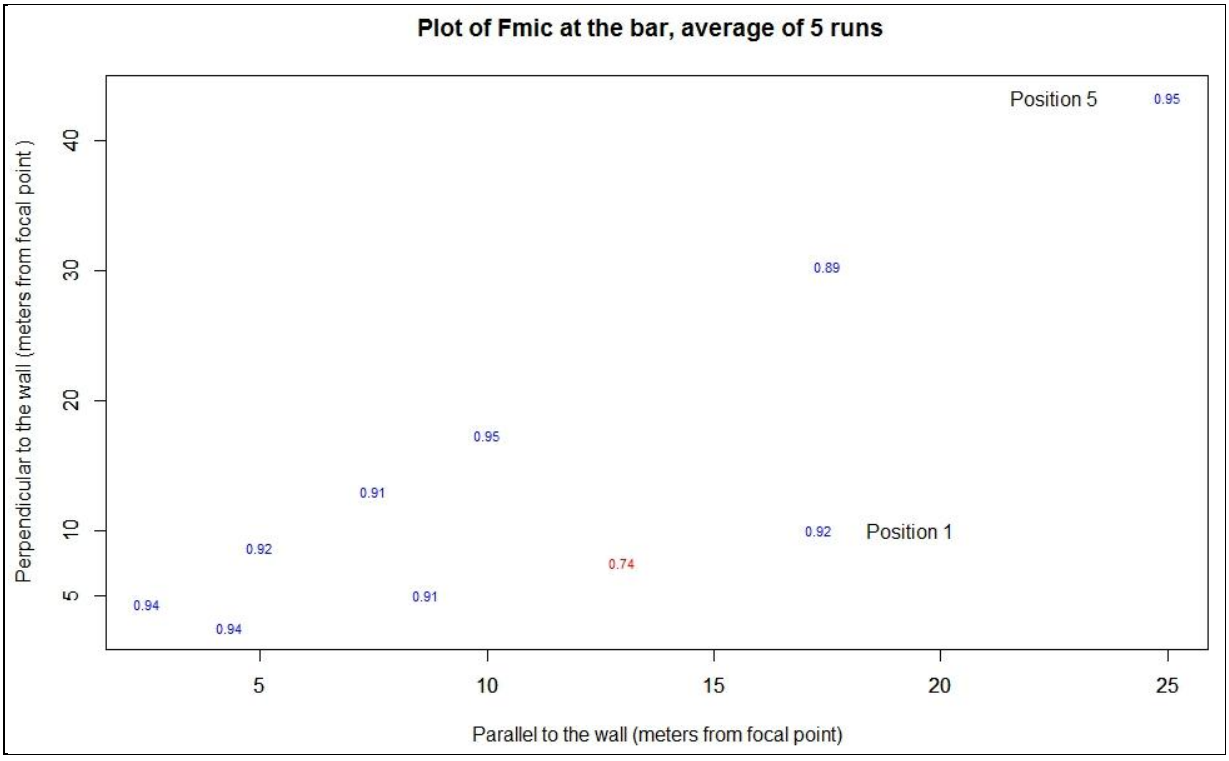

Figure A-3. Micro *F* measure around the bar.

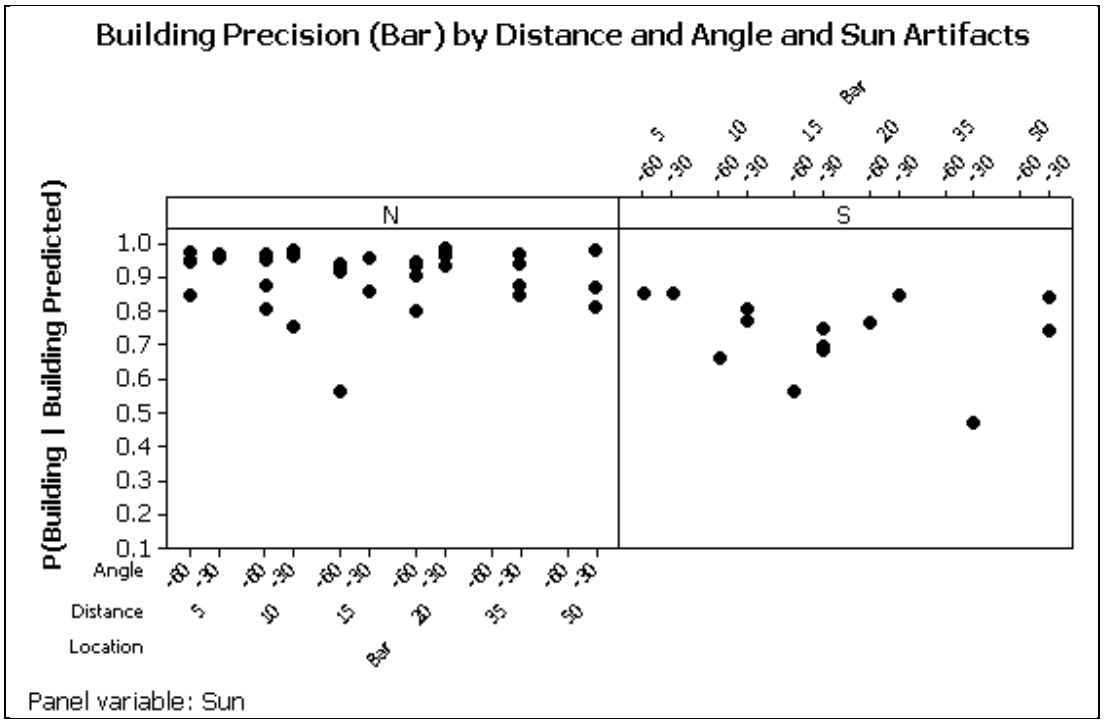

Figure A-4. Bar precision by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

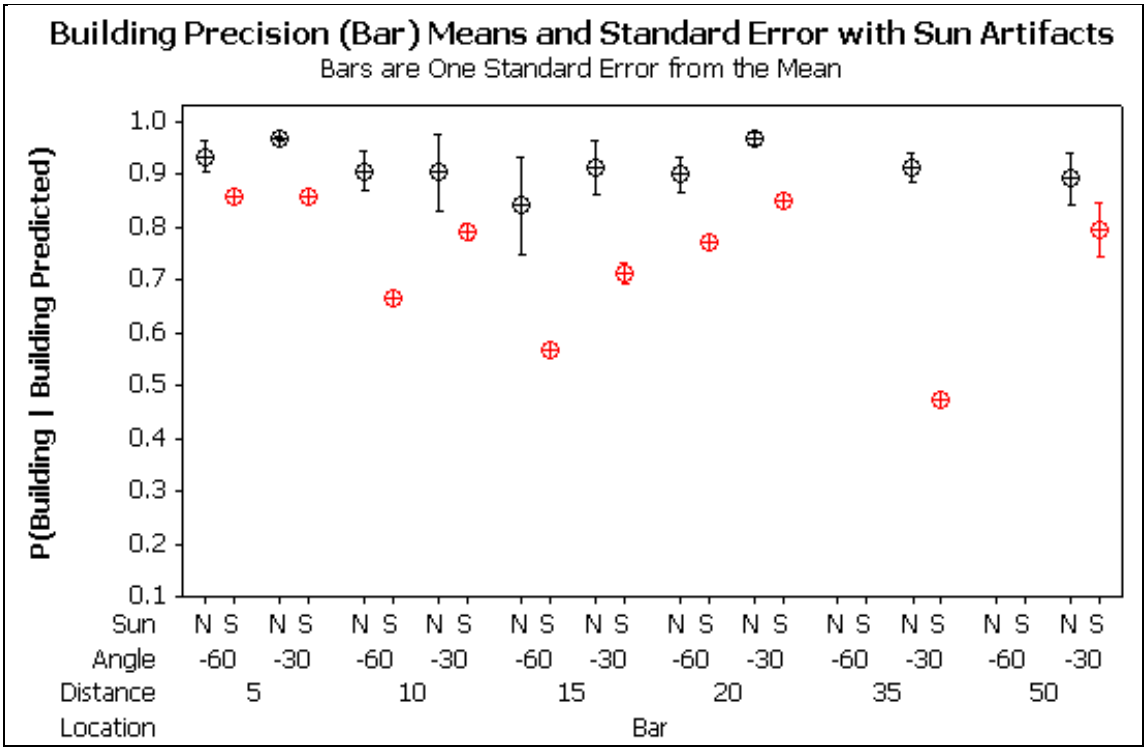

Figure A-5. Bar precision plot of means (circle) with bars, indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

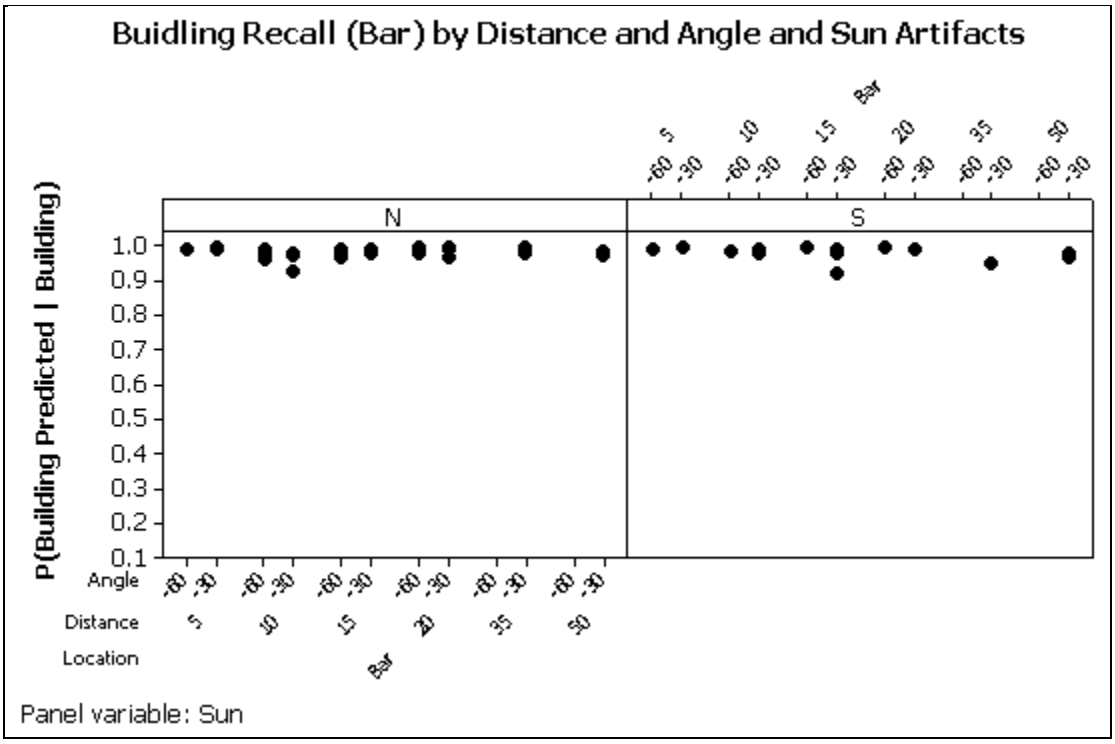

Figure A-6. Bar recall by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

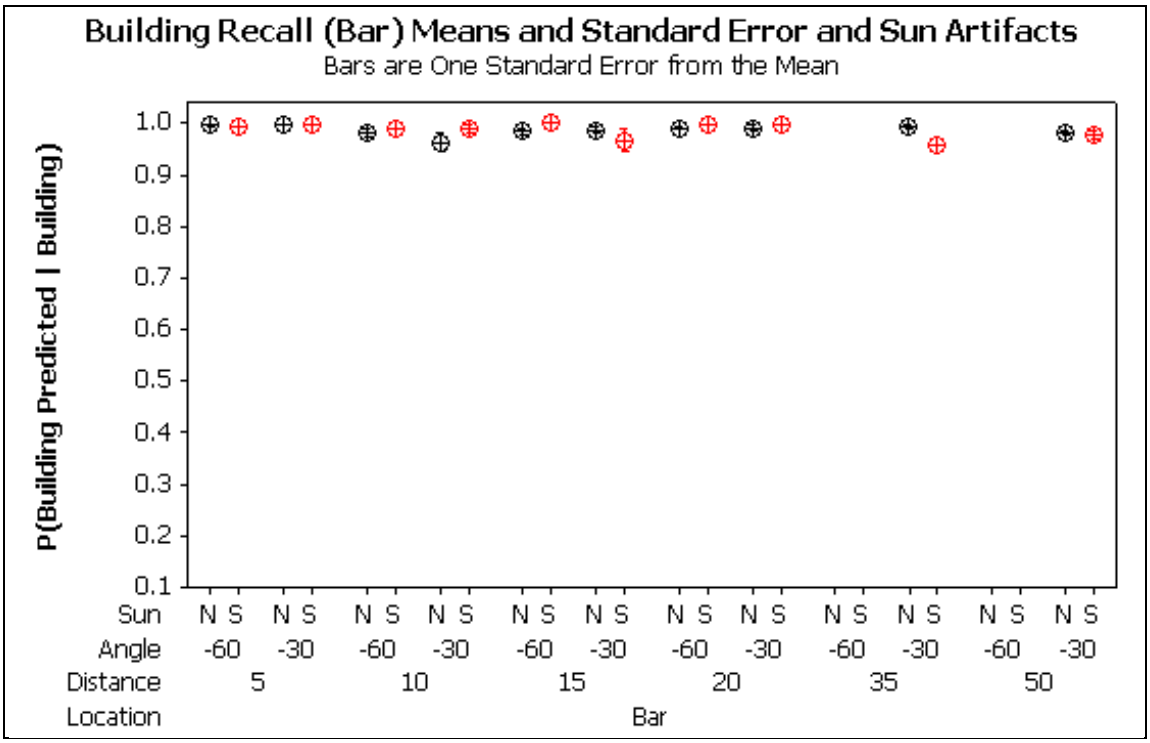

Figure A-7. Bar recall plot of means (circle) with bars, indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

Table A-2. Bar location ANOVA for building precision under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BP)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.38428	0.37144	0.18572	13.29	0.000
Distance	5	0.25521	0.31829	0.06366	4.55	0.003
Angle	1	0.01778	0.03278	0.03278	2.34	0.137
Time*distance	10	0.20279	0.20522	0.02052	1.47	0.202
Time*angle	2	0.19309	0.19309	0.09654	6.91	0.004
Error	29	0.40541	0.40541	0.01398	—	—
Total	49	1.45855	—	—	—	—

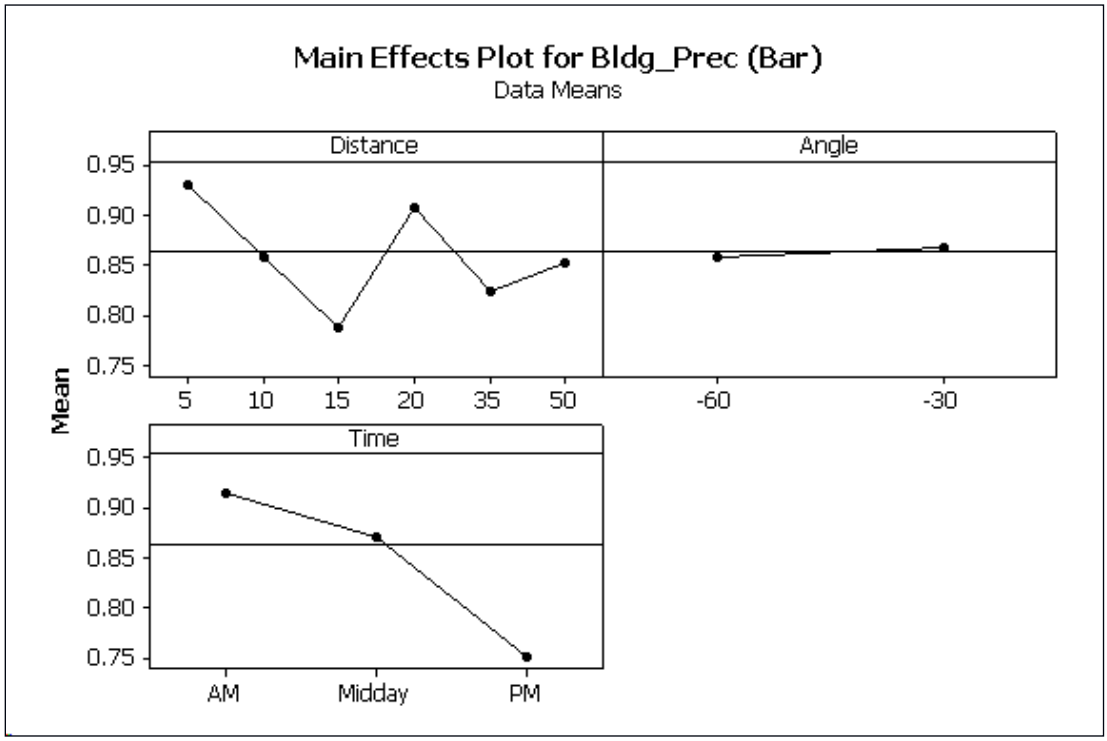

Figure A-8. Bar precision ANOVA main effects.

Table A-3. Bar location ANOVA for building recall under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BR)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.002215	0.001637	0.000819	0.35	0.709
Distance	5	0.043068	0.033041	0.006608	2.81	0.034
Angle	1	0.002133	0.005607	0.005607	2.39	0.133
Time*distance	10	0.025904	0.019262	0.001926	0.82	0.613
Time*angle	2	0.010711	0.010711	0.005356	2.28	0.120
Error	29	0.068157	0.068157	0.002350	—	—
Total	49	0.152187	—	—	—	—

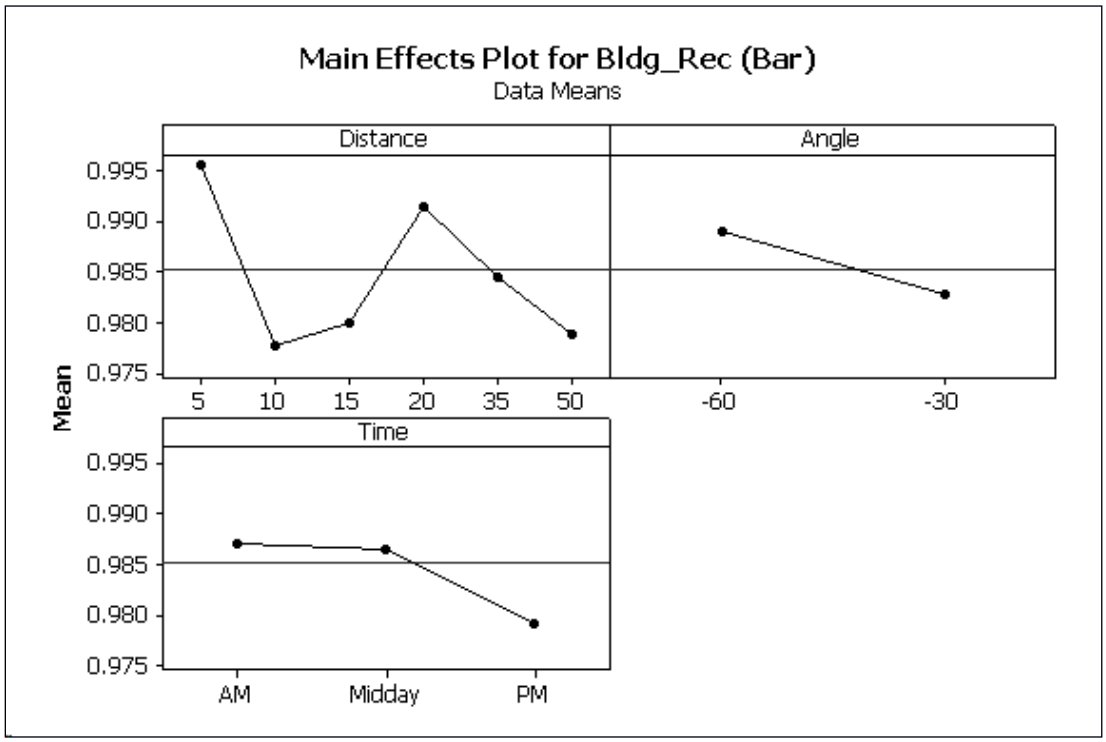

Figure A-9. Bar recall ANOVA main effects.

INTENTIONALLY LEFT BLANK.

Appendix B. The Church

Figures B-1 through B-9 and tables B-1 through B-3 show details and results for the church building.

Figure B-1. The church.

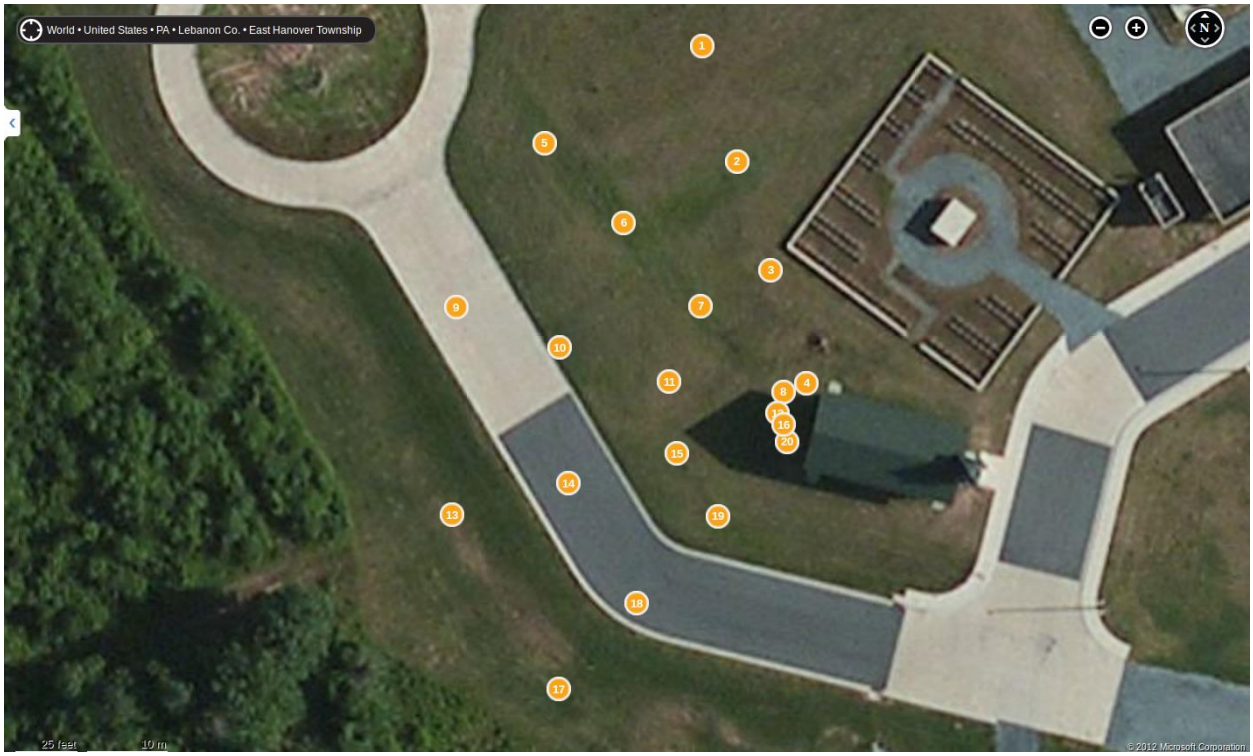

Figure B-2. Data collection positions around the church.

Table B-1. *F* measure by distances and angle around the church.

Position	Angle (°)	Dist (m)	Fmac	Fmic	P Bldg	R Bldg
1	60	50	0.657	0.835	0.425	0.924
2	60	35	0.585	0.840	0.587	0.985
3	60	20	0.747	0.925	0.787	0.935
4	60	5	0.628	0.869	0.903	0.918
5	30	50	0.889	0.952	0.630	0.870
6	30	35	0.897	0.946	0.662	0.870
7	30	20	0.459	0.599	0.512	0.840
8	30	5	0.645	0.872	0.973	0.828
9	0	50	0.817	0.968	0.637	0.821
10	0	35	0.852	0.943	0.632	0.924
11	0	20	0.879	0.903	0.667	0.940
12	0	5	0.832	0.902	0.977	0.837
13	–30	50	0.649	0.879	0.570	0.812
14	–30	35	0.833	0.962	0.795	0.944
15	–30	20	0.881	0.969	0.854	0.959
16	–30	5	0.724	0.947	0.959	0.946
17	–60	50	0.864	0.968	0.707	0.901
18	–60	35	0.847	0.973	0.871	0.959
19	–60	20	0.912	0.951	0.878	0.878
20	–60	5	0.740	0.950	0.935	0.970

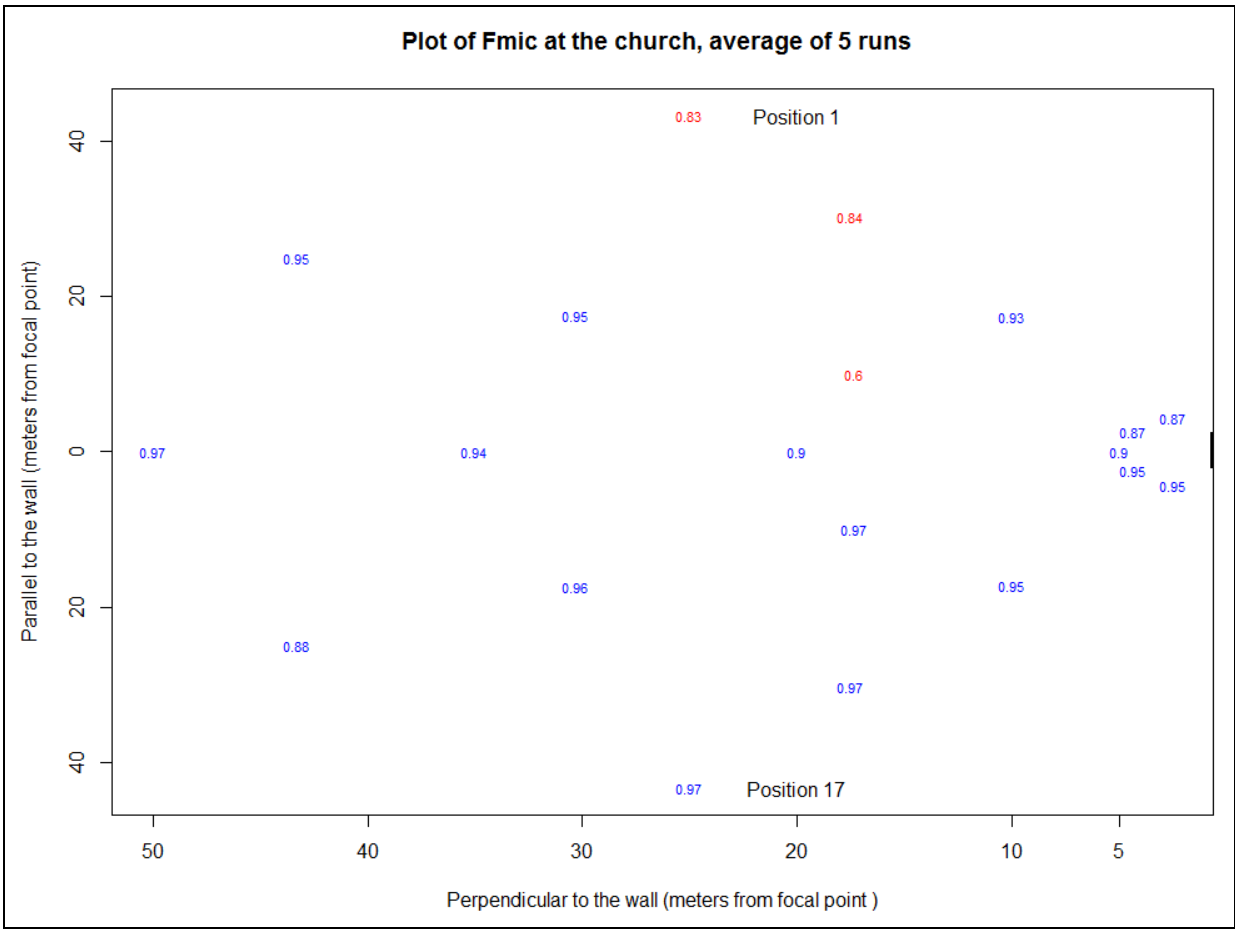

Figure B-3. Micro *F* measure around the church.

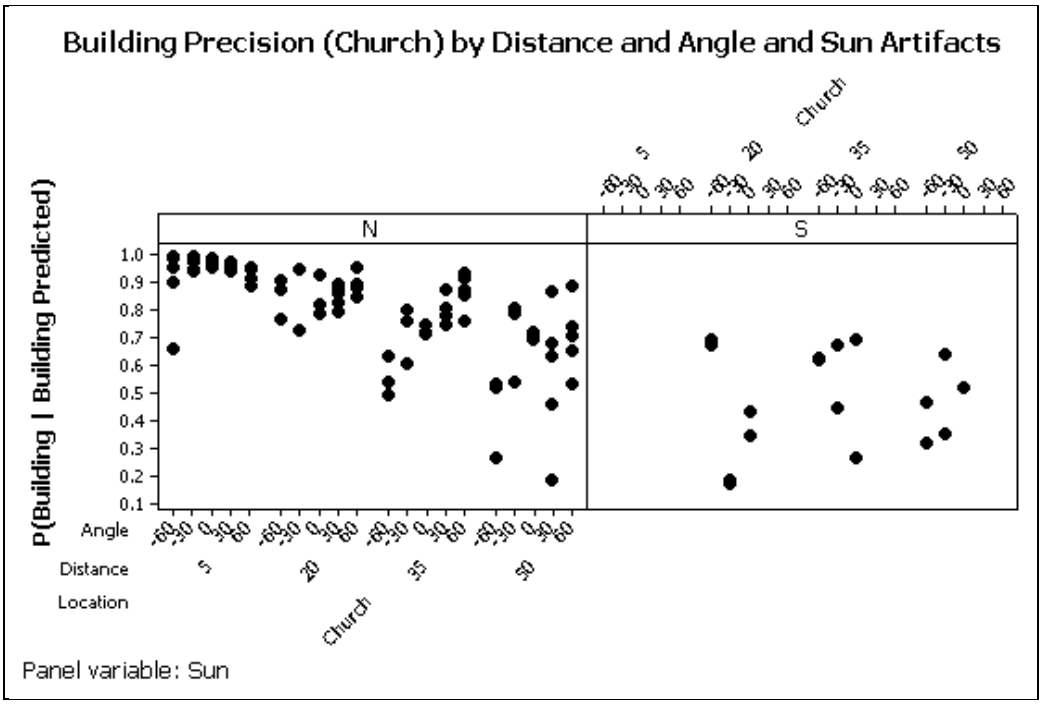

Figure B-4. Church precision by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

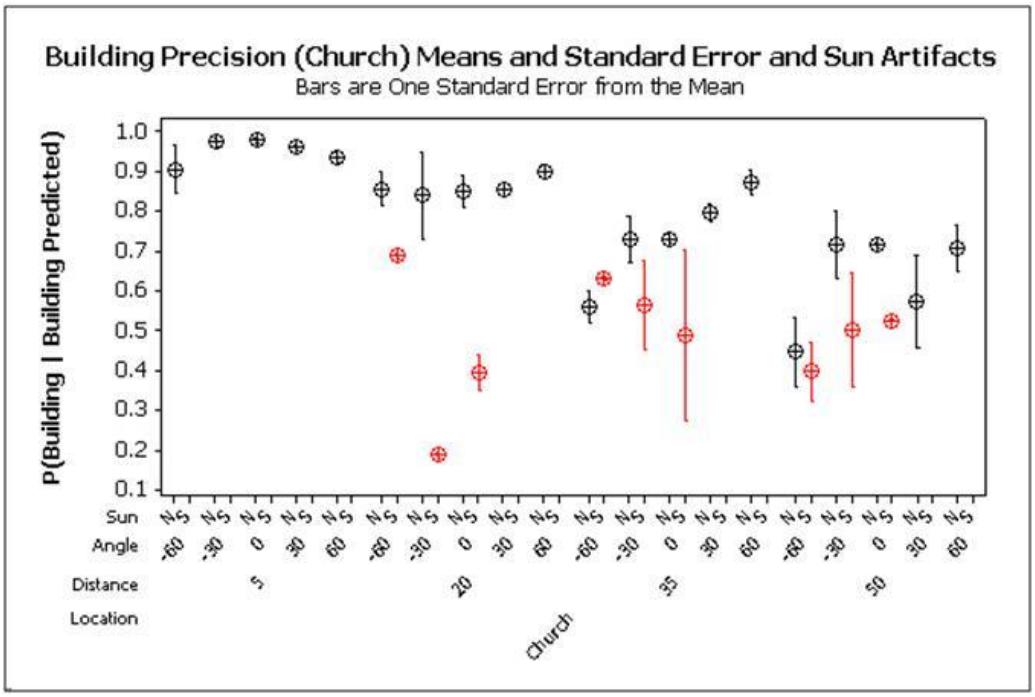

Figure B-5. Church precision plot of means (circle) with bars indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

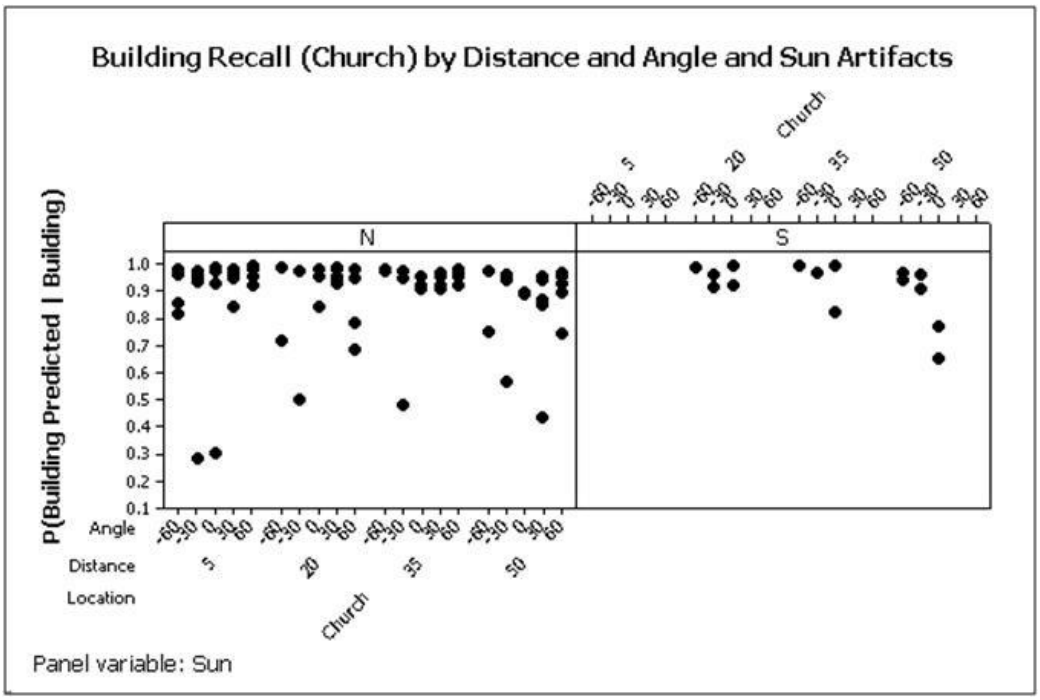

Figure B-6. Church recall by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

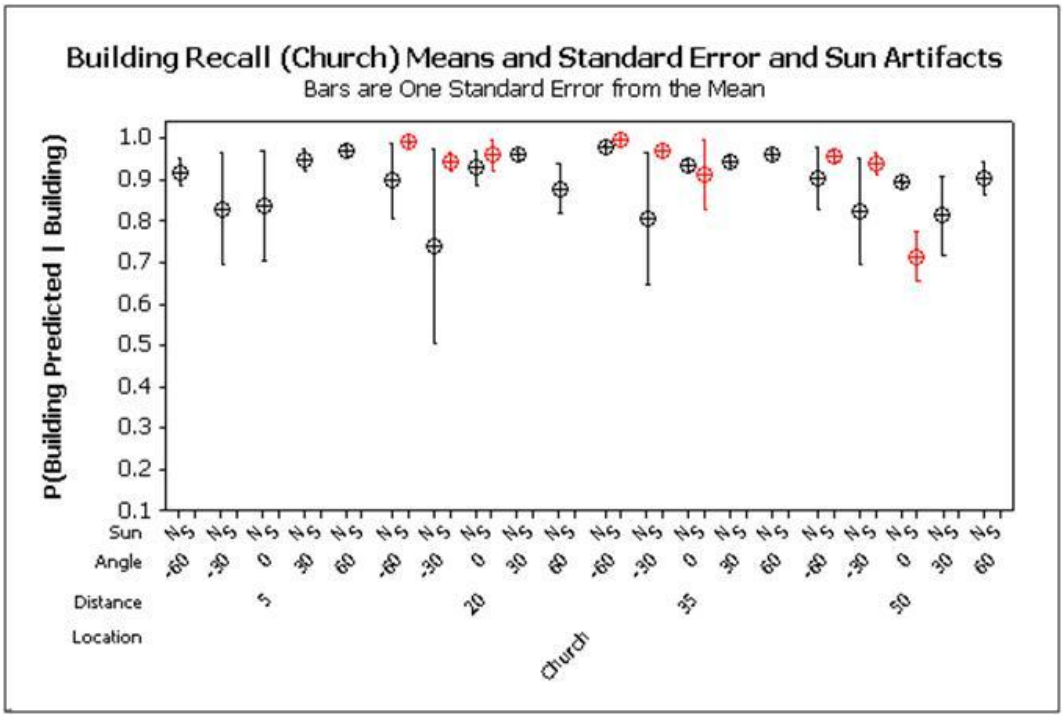

Figure B-7. Church recall plot of means (circle) with bars, indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

Table B-2. Church location ANOVA for building precision under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BP)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.41261	0.43266	0.21633	13.62	0.000
Distance	3	3.14589	2.91216	0.97072	61.10	0.000
Angle	4	0.48394	0.44506	0.11126	7.00	0.000
Time*distance	6	0.40554	0.42530	0.07088	4.46	0.002
Time*angle	8	0.40836	0.40728	0.05091	3.20	0.007
Distance*angle	12	0.65897	0.51041	0.04253	2.68	0.010
Time*distance*angle	24	0.55933	0.55933	0.02331	1.47	0.141
Error	39	0.61963	0.61963	0.01589	—	—
Total	98	6.69428	—	—	—	—

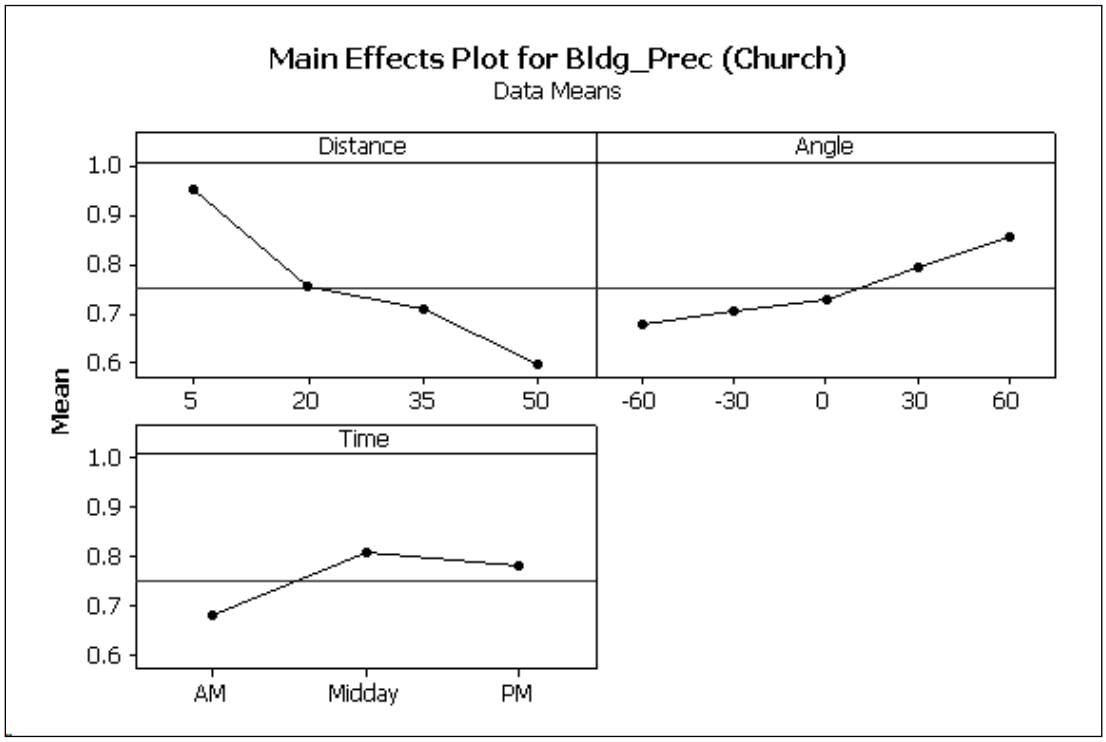

Figure B-8. Church precision ANOVA main effects.

Table B-3. Church location ANOVA for building recall under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BR)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.32066	0.32305	0.16153	5.69	0.007
Distance	3	0.19955	0.26146	0.08715	3.07	0.039
Angle	4	0.19957	0.23241	0.05810	2.05	0.107
Time*distance	6	0.64324	0.64212	0.10702	3.77	0.005
Time*angle	8	0.36727	0.36600	0.04575	1.61	0.153
Distance*angle	12	0.22977	0.25813	0.02151	0.76	0.688
Time*distance*angle	24	0.64010	0.64010	0.02667	0.94	0.555
Error	39	1.10707	1.10707	0.02839	—	—
Total	98	3.70722	—	—	—	—

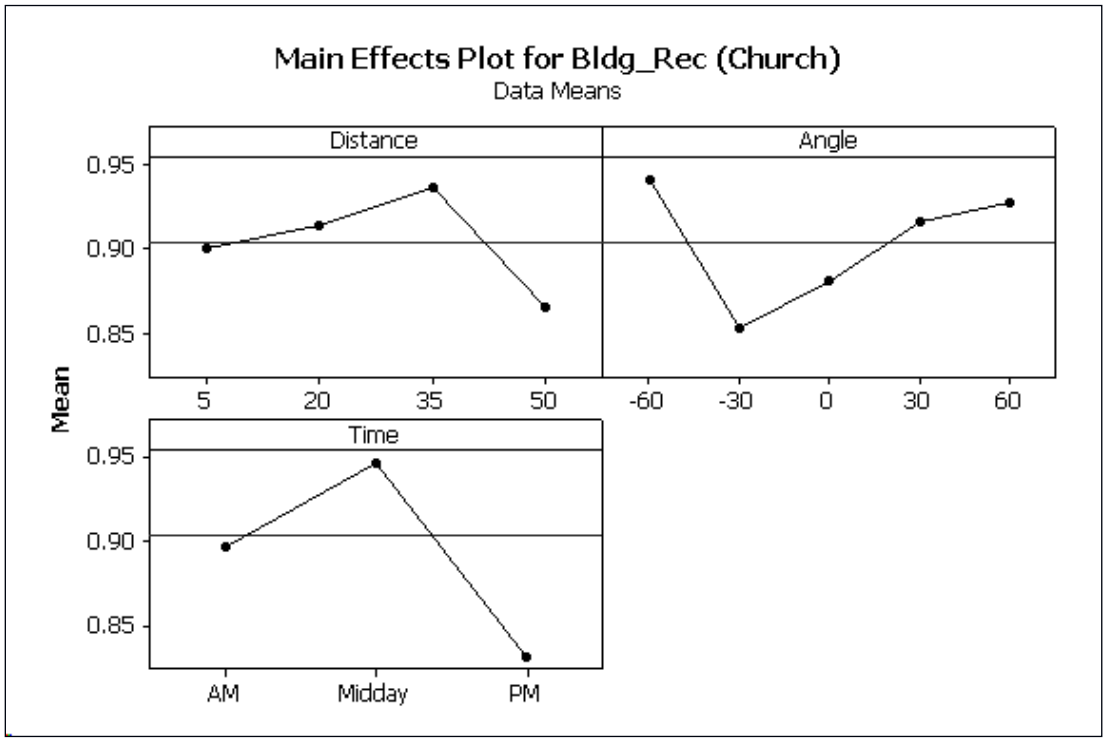

Figure B-9. Church recall ANOVA main effects.

Appendix C. The Cube

Figures C-1 through C-9 and tables C-1 through C-3 show details and results for the cube building.

Figure C-1. The cube is the foreground building with the gray door.

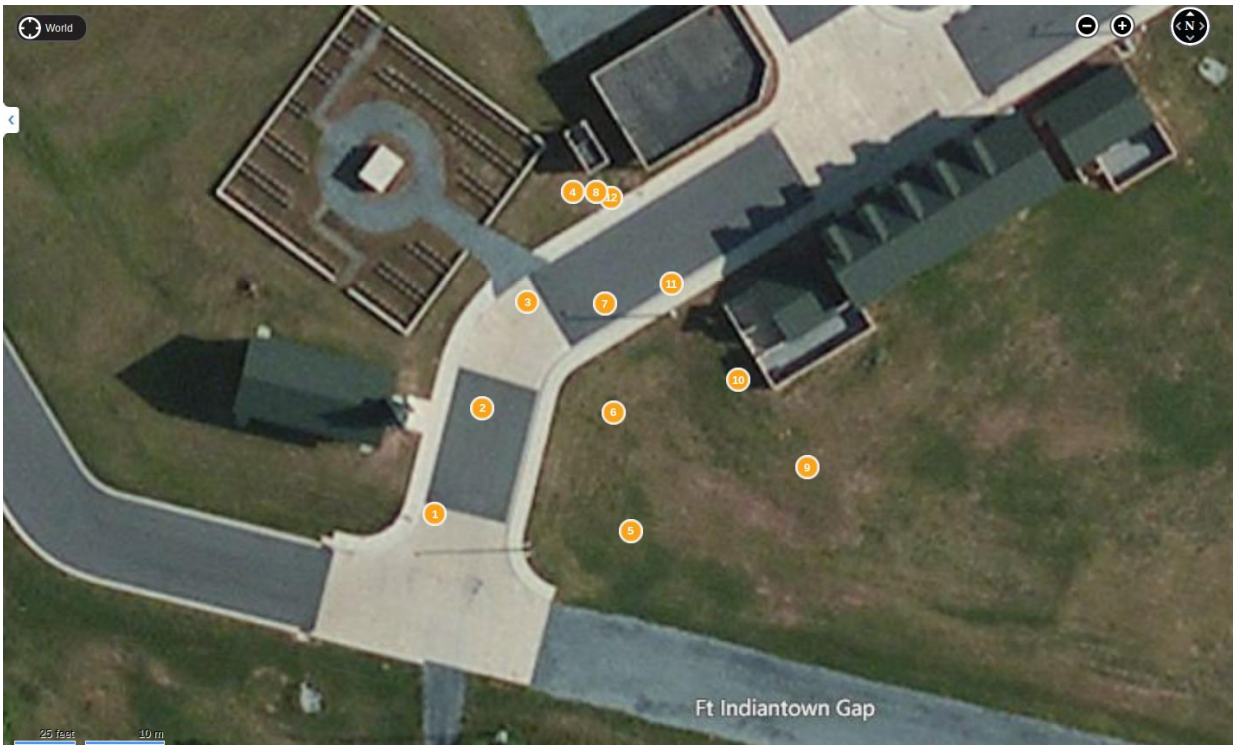

Figure C-2. Data collection positions around the cube.

Table C-1. *F* measure by distances and angle around the cube.

Position	Angle (°)	Dist (m)	Fmac	Fmic	P Bldg	R Bldg
1	30	50	0.632	0.883	0.628	0.948
2	30	35	0.716	0.940	0.519	0.953
3	30	20	0.821	0.911	0.698	0.921
4	30	5	0.853	0.958	0.954	0.950
5	0	50	0.937	0.964	0.882	0.940
6	0	35	0.761	0.969	0.654	0.897
7	0	20	0.872	0.958	0.792	0.981
8	0	5	0.896	0.928	0.941	0.908
9	–30	50	0.776	0.978	0.904	0.965
10	–30	35	0.763	0.969	0.851	0.970
11	–30	20	0.841	0.958	0.786	0.979
12	–30	5	0.943	0.948	0.894	0.974

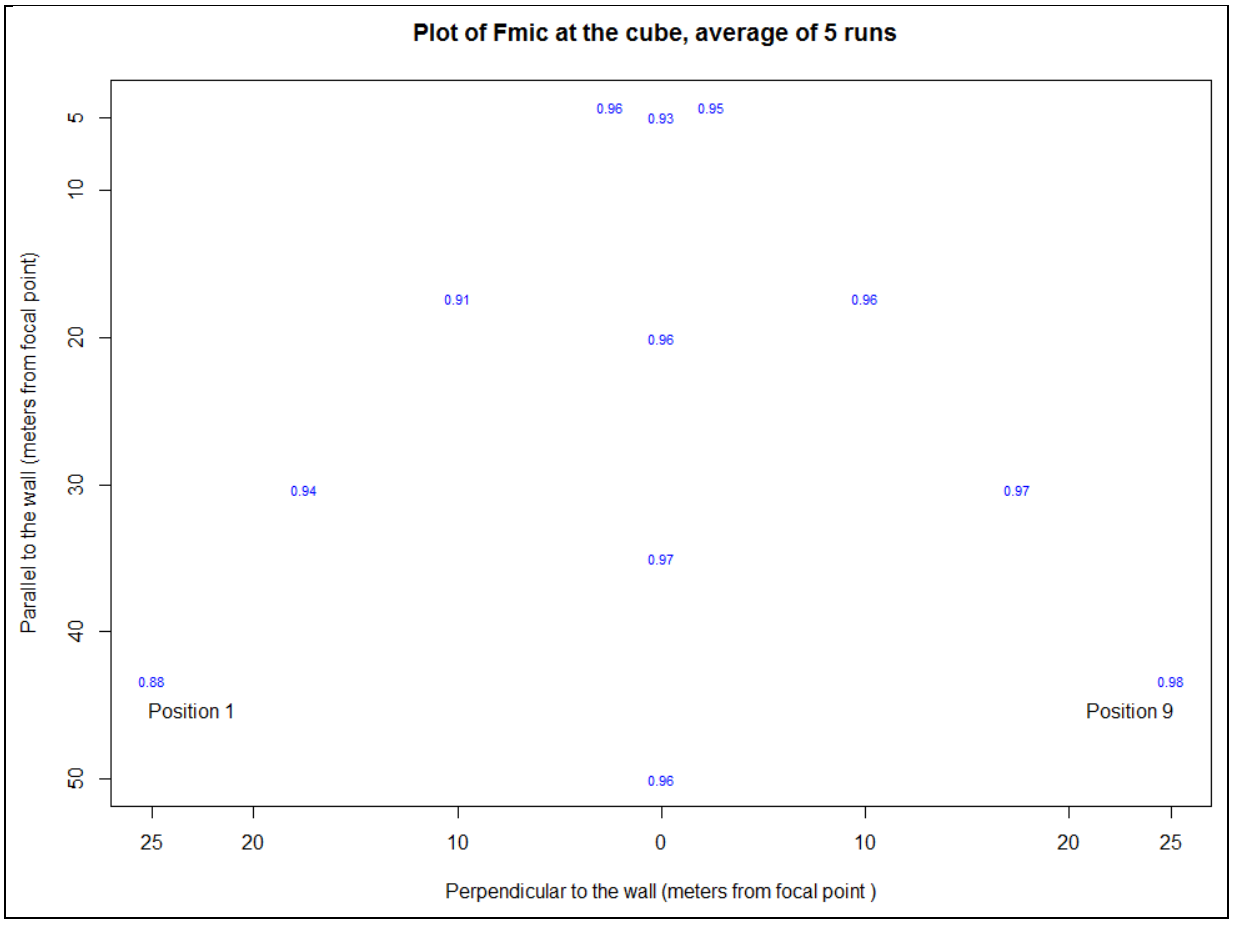

Figure C-3. Micro *F* measure around the cube.

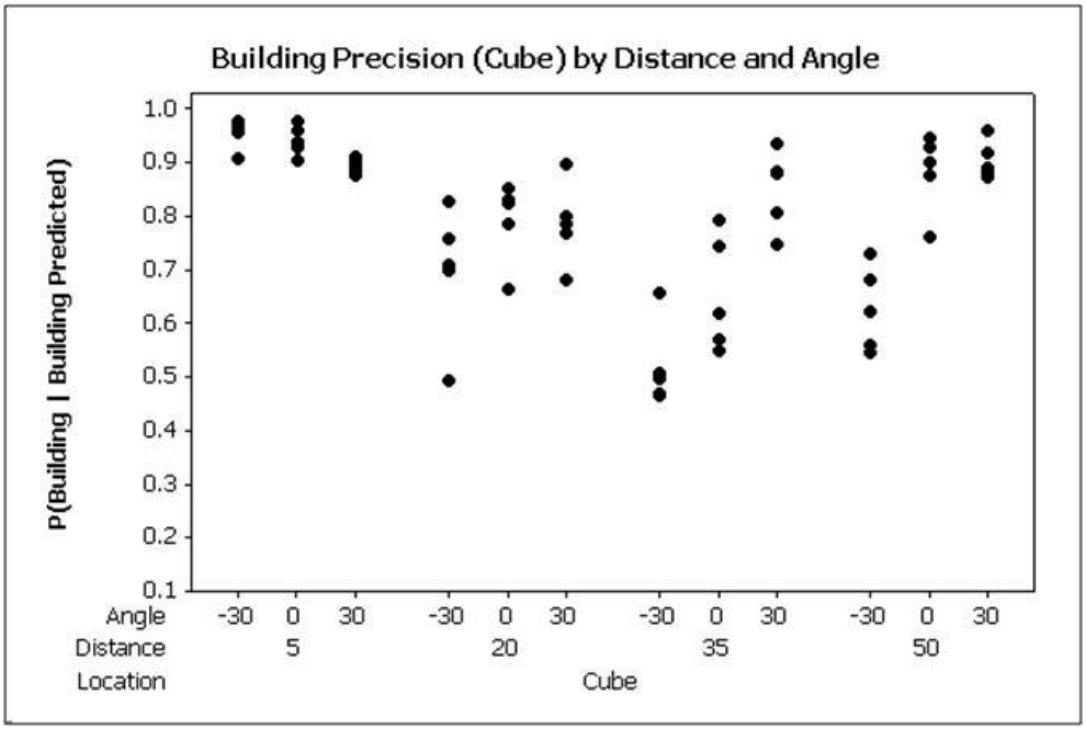

Figure C-4. Cube precision by distance and angle. There were no images with sun interference.

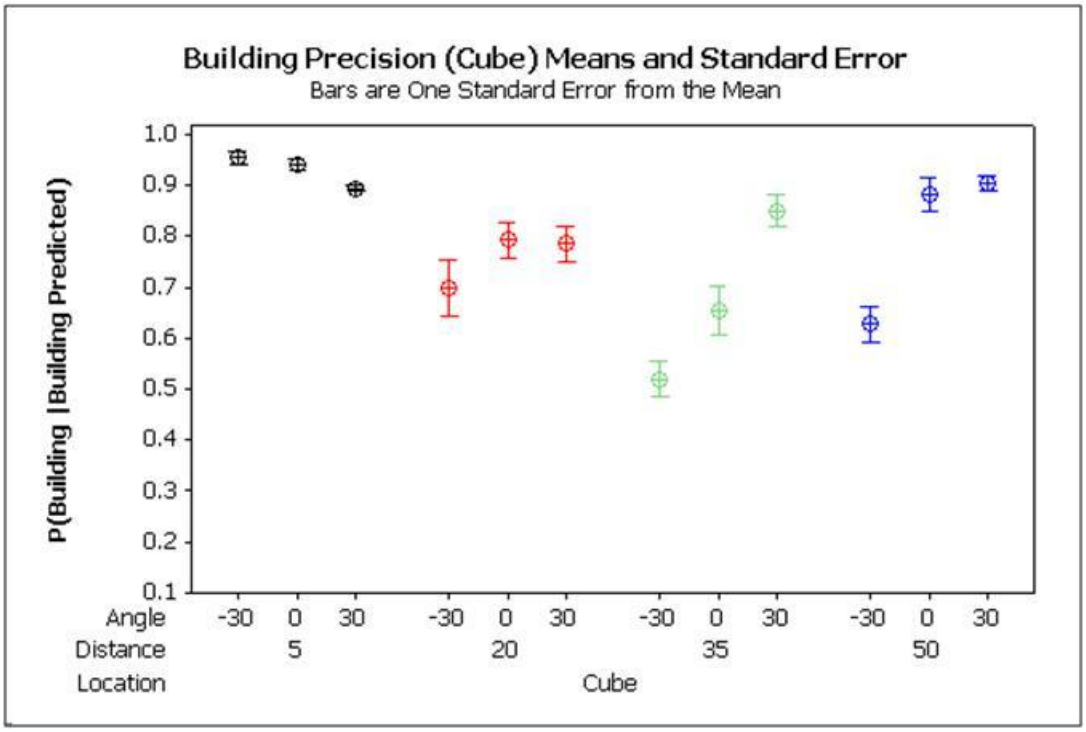

Figure C-5. Cube precision plot of means (circle) with bars indicating one standard error from the mean.

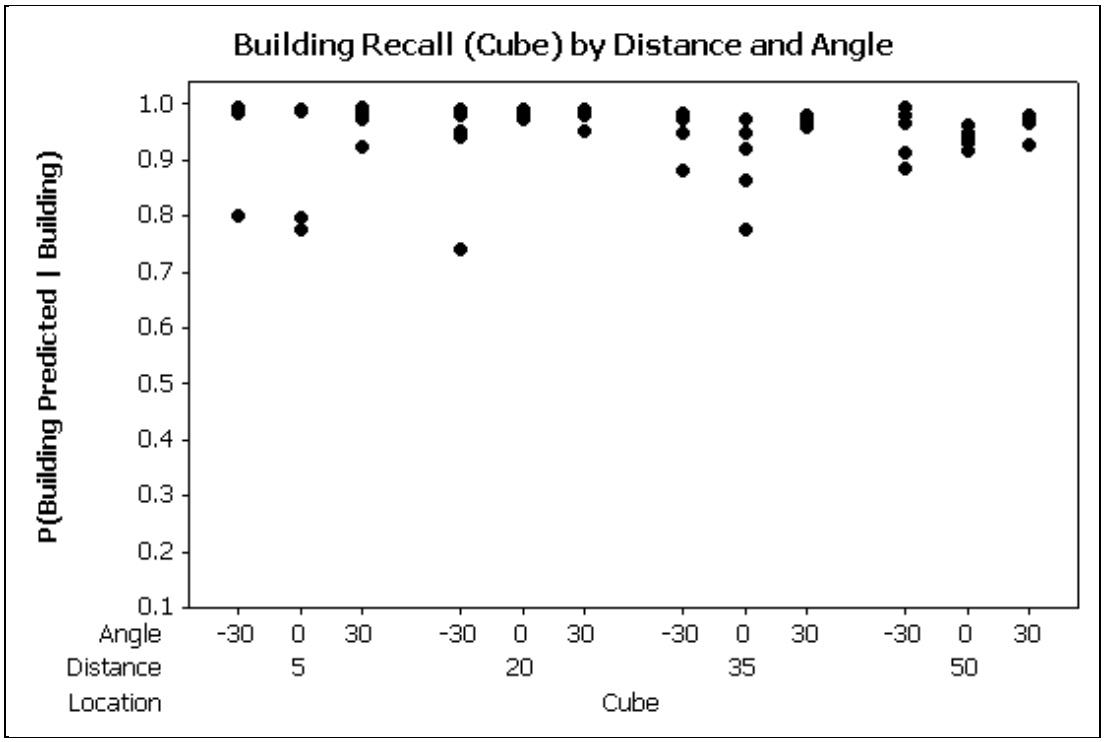

Figure C-6. Cube recall by distance and angle.

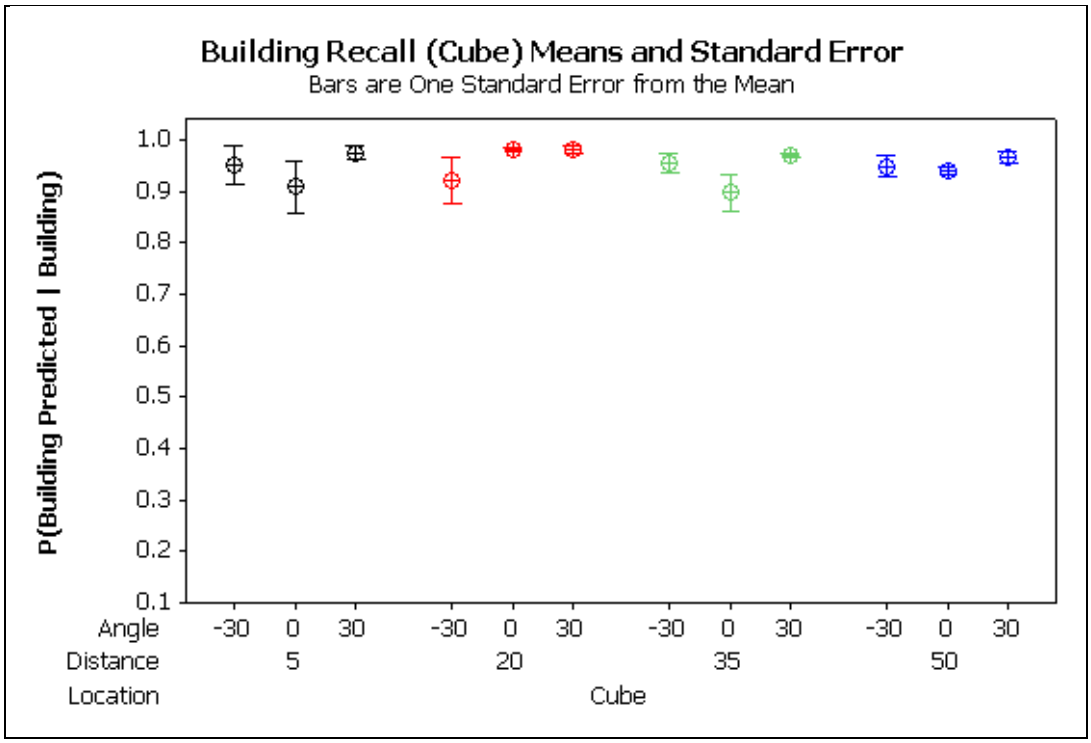

Figure C-7. Cube recall plot of means (circle) with bars, indicating one standard error from the mean.

Table C-2. Cube location ANOVA for building precision under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BP)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.007005	0.007005	0.003503	0.36	0.701
Distance	3	0.897293	0.847860	0.282620	29.04	0.000
Angle	2	0.336212	0.326808	0.163404	16.79	0.000
Time*distance	6	0.032675	0.032675	0.005446	0.56	0.758
Time*angle	4	0.042857	0.042857	0.010714	1.10	0.379
Distance*angle	6	0.468332	0.447373	0.074562	7.66	0.000
Time*distance*angle	12	0.068261	0.068261	0.005688	0.58	0.833
Error	24	0.233582	0.233582	0.009733	—	—
Total	59	2.086217	—	—	—	—

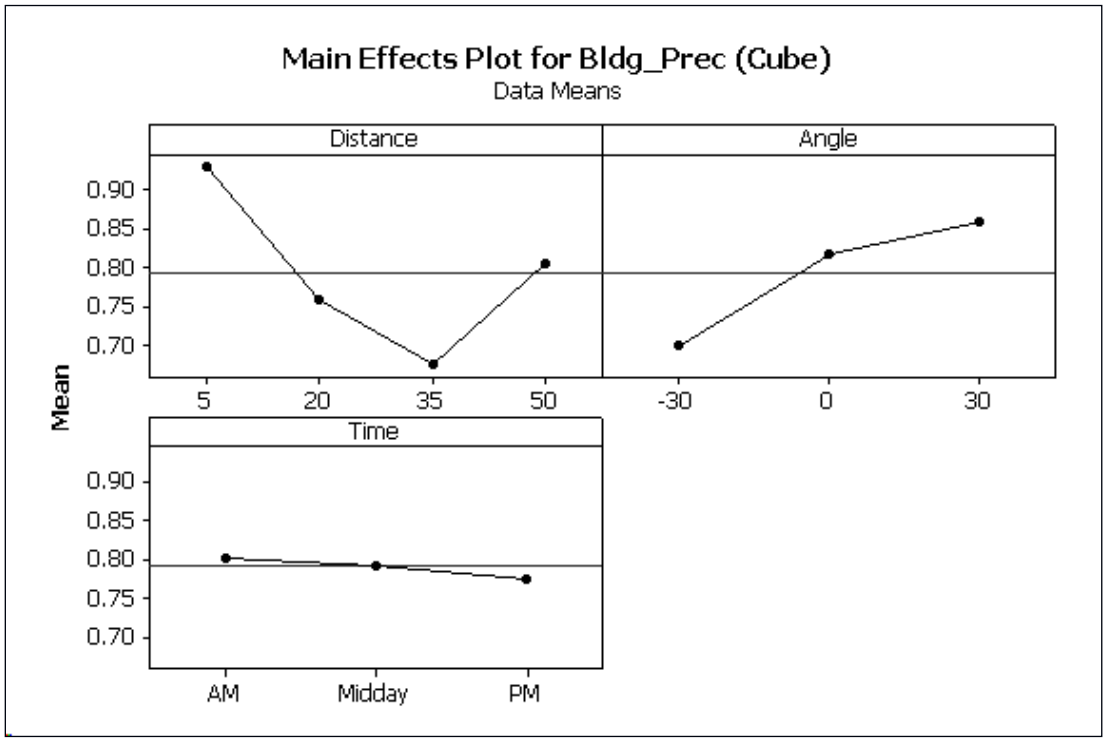

Figure C-8. Cube precision ANOVA main effects.

Table C-3. Cube location ANOVA for building recall under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BR)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.121261	0.121261	0.060630	7.13	0.004
Distance	3	0.024348	0.007354	0.002451	0.29	0.833
Angle	2	0.061772	0.046608	0.023304	2.74	0.085
Time*distance	6	0.078197	0.078197	0.013033	1.53	0.210
Time*angle	4	0.032699	0.032699	0.008175	0.96	0.447
Distance*angle	6	0.070546	0.119496	0.019916	2.34	0.064
Time*distance*angle	12	0.156893	0.156893	0.013074	1.54	0.178
Error	24	0.204103	0.204103	0.008504	—	—
Total	59	0.749819	—	—	—	—

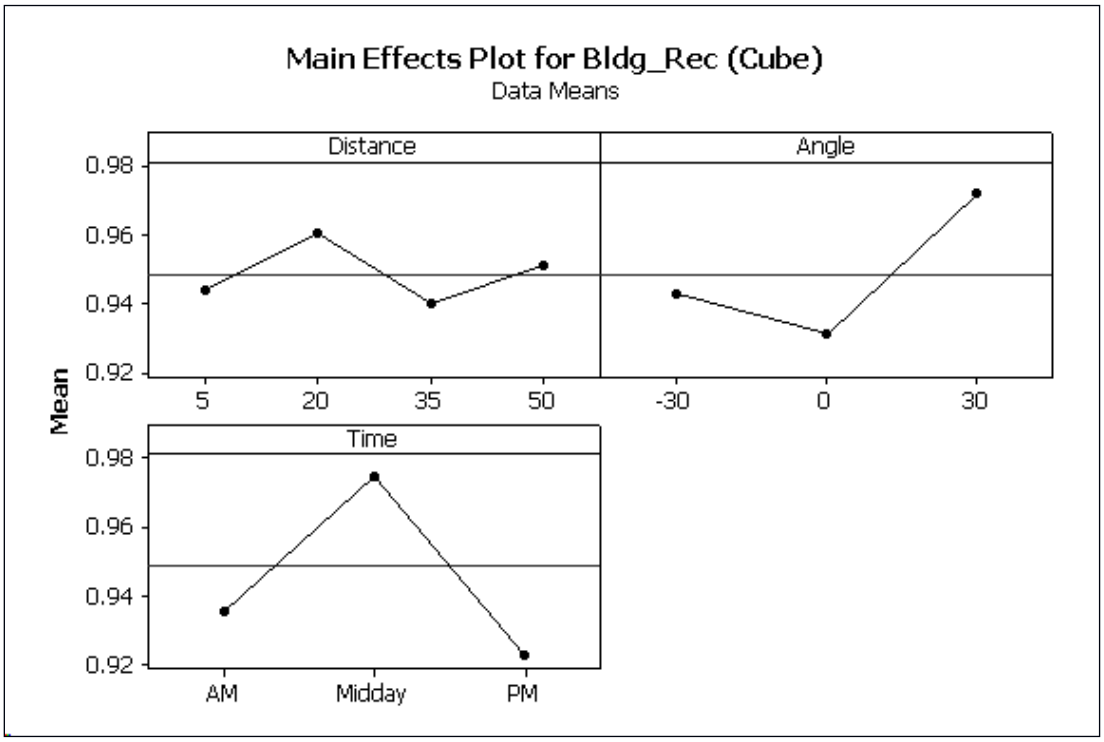

Figure C-9. Cube recall ANOVA main effects.

INTENTIONALLY LEFT BLANK.

Appendix D. The Police Station

Figures D-1 through D-9 and tables D-1 through D-3 show details and results for the police station building.

Figure D-1. The police station.

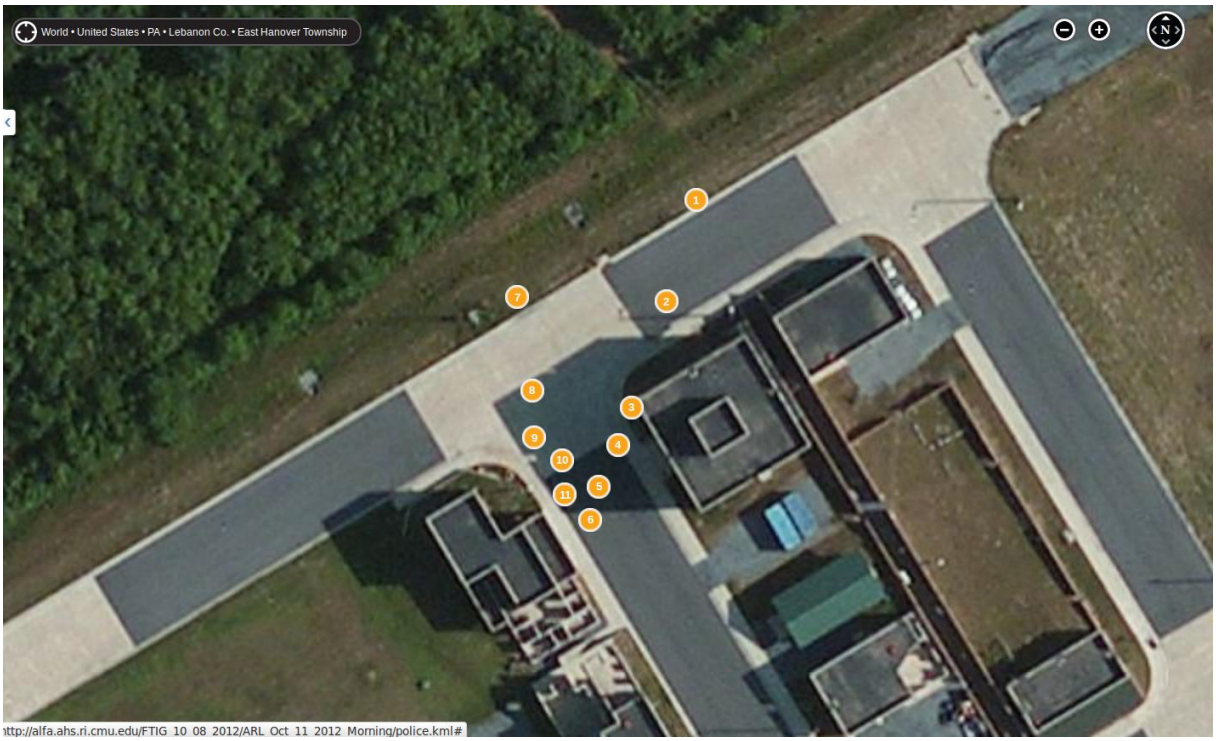

Figure D-2. Data collection positions around the police station.

Table D-1. *F* measure by distances and angle around the police station.

Position	Angle (°)	Dist (m)	Fmac	Fmic	P Bldg	R Bldg
1	–30	50	0.889	0.929	0.89	0.987
2	–30	35	0.748	0.913	0.899	0.97
3	–30	20	0.674	0.918	0.839	0.988
4	–30	15	0.796	0.954	0.897	0.984
5	–30	10	0.658	0.899	0.828	0.967
6	–30	5	0.897	0.974	0.981	0.985
7	–60	35	0.557	0.818	0.753	0.93
8	–60	20	0.667	0.912	0.803	0.978
9	–60	15	0.703	0.898	0.835	0.983
10	–60	10	0.719	0.829	0.643	0.967
11	–60	5	0.653	0.883	0.887	0.914

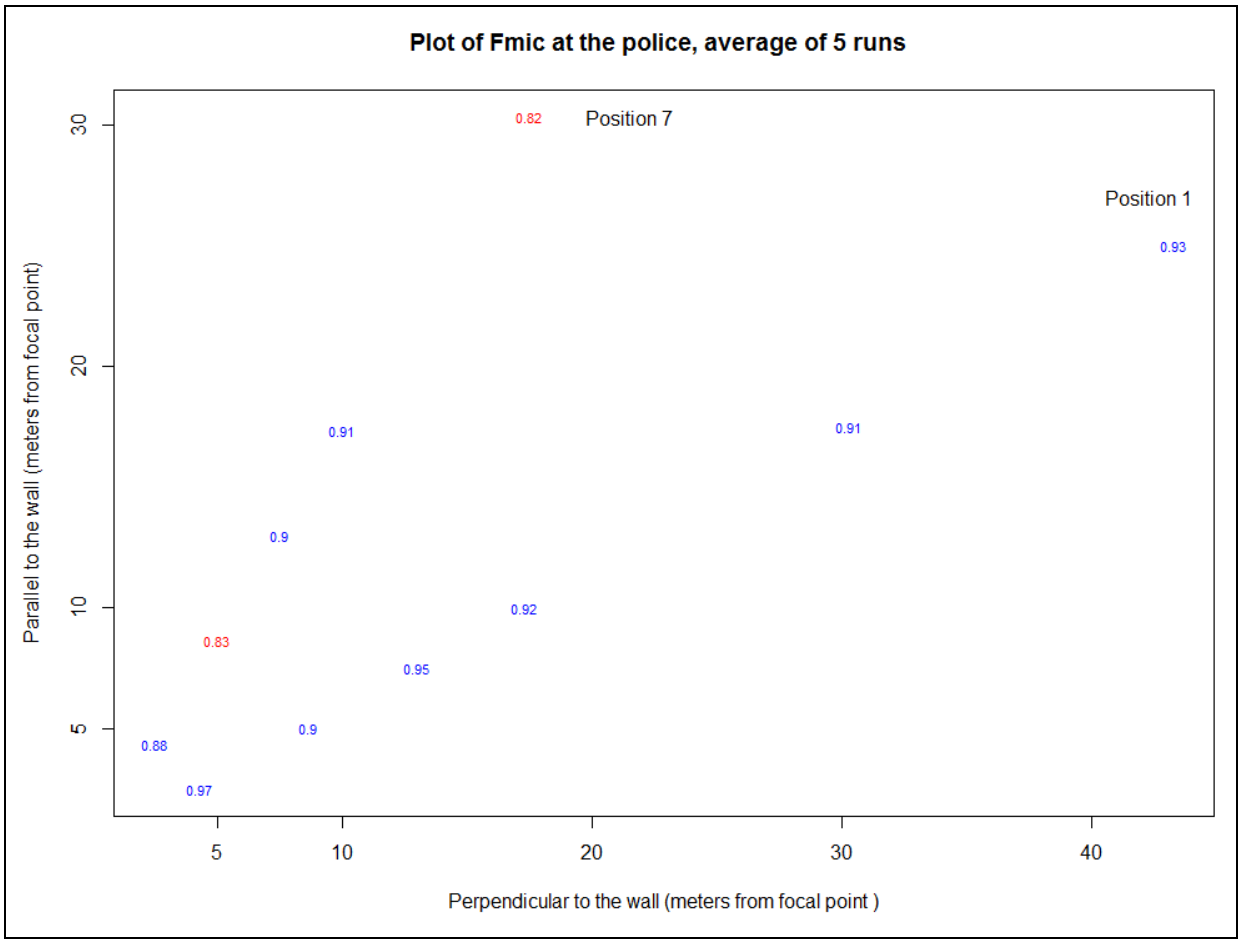

Figure D-3. Micro *F* measure around the police station.

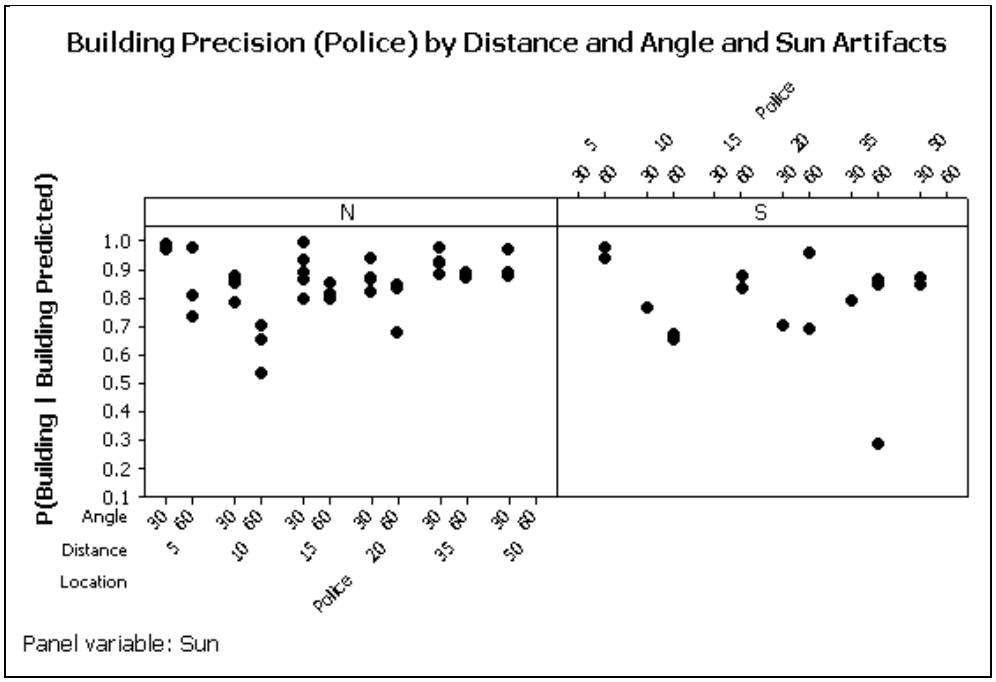

Figure D-4. Police station precision by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

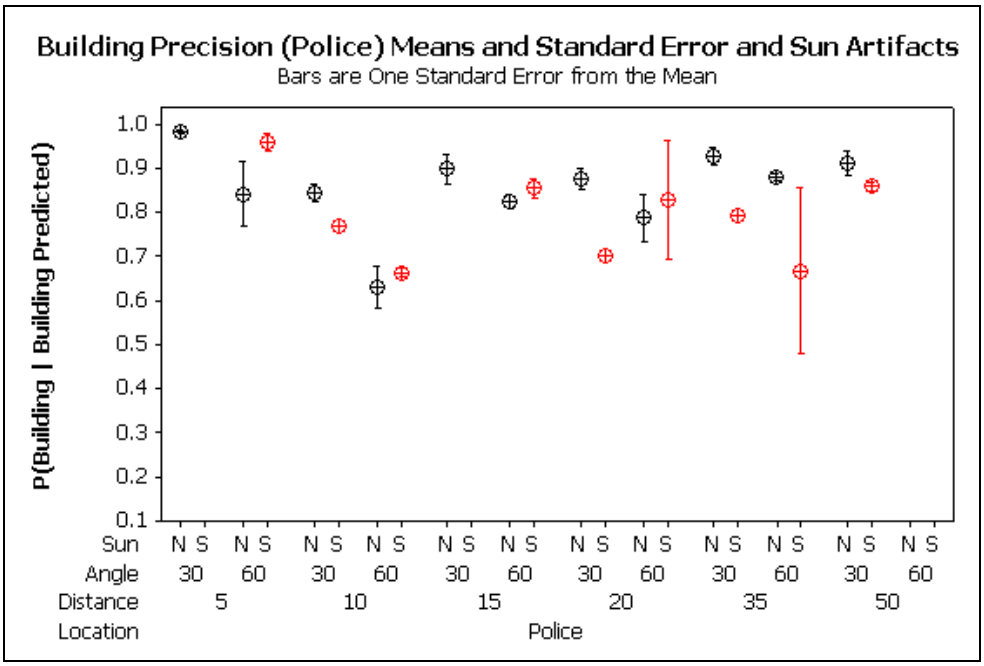

Figure D-5. Police station precision plot of means (circle) with bars, indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

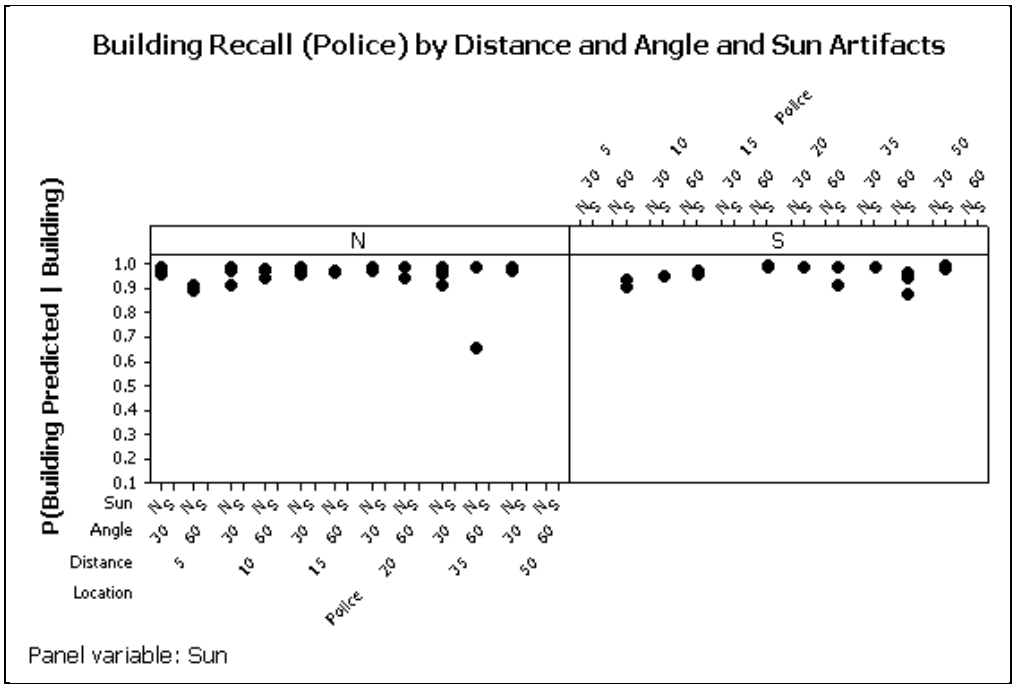

Figure D-6. Police station recall by distance and angle.

Note: The left panel shows the results with no sun interference, and the right panel shows results with sun interference.

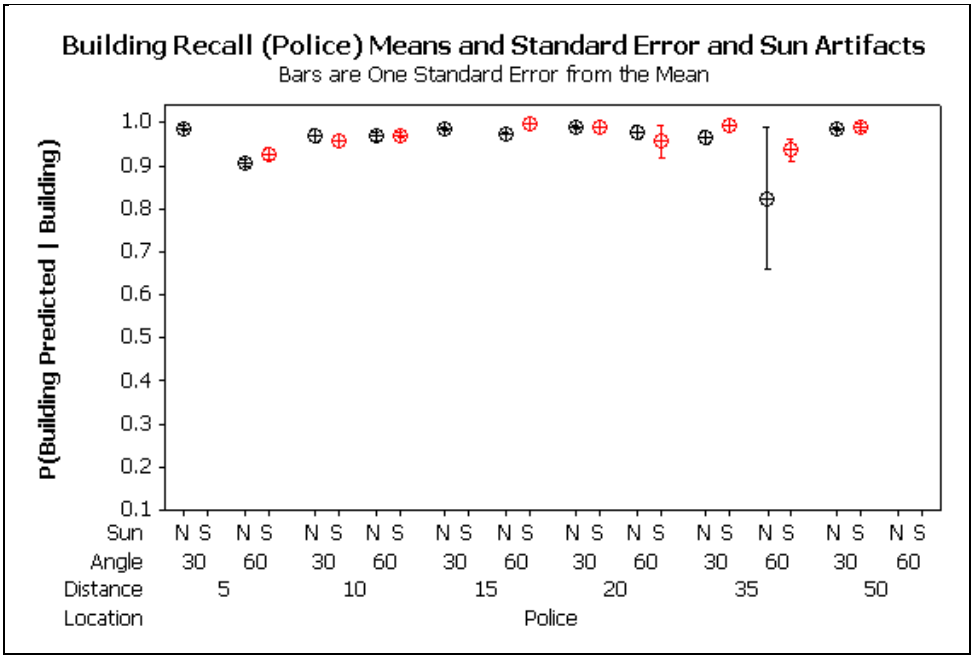

Figure D-7. Police station recall plot of means (circle) with bars, indicating one standard error from the mean.

Note: Red circles represent images with sun interference, and black circles represent those with no sun interference.

Table D-2. Police location ANOVA for building precision under $\mathrm{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BP)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.16933	0.16933	0.08467	5.42	0.008
Distance	5	0.51607	0.50241	0.10048	6.43	0.000
Angle	1	0.27050	0.27050	0.27050	17.32	0.000
Error	46	0.71830	0.71830	0.01562	—	—
Total	54	1.67421	—	—	—	—

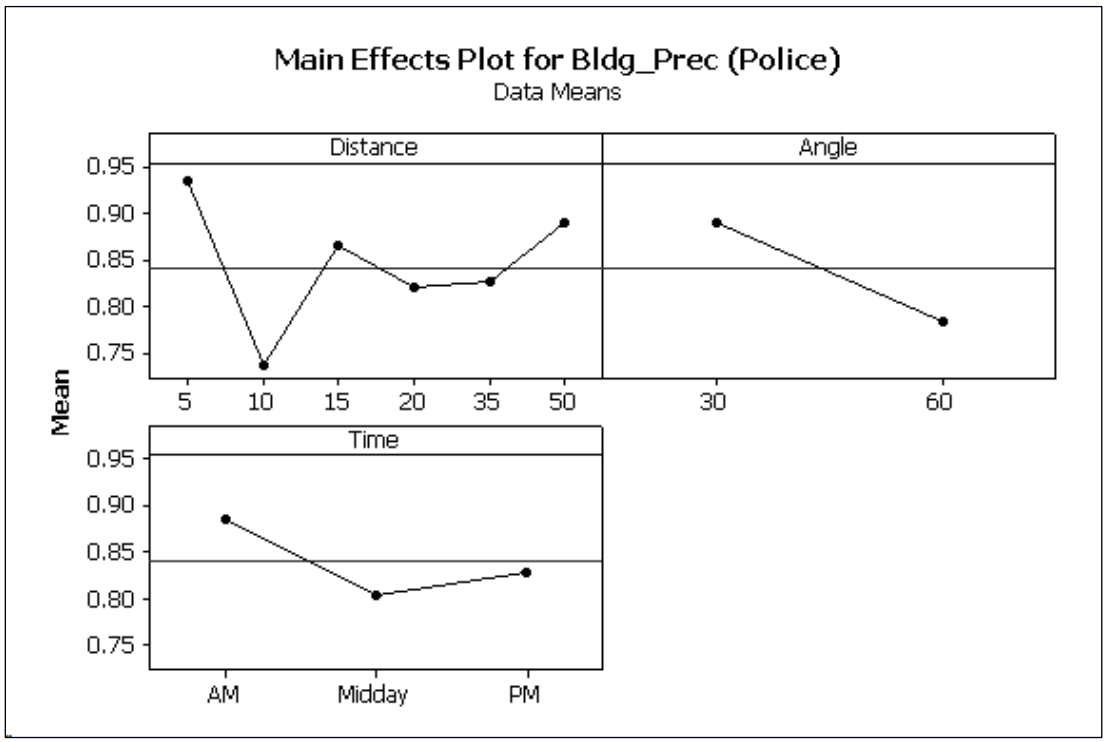

Figure D-8. Police precision ANOVA main effects.

Table D-3. Police location ANOVA for building recall under $\text{Sin}^{-1}(\sqrt{p})$ transformation (ASSq(BR)).

Source	DF	Seq SS	Adj SS	Adj MS	*F*	P
Time	2	0.019934	0.019934	0.009967	1.28	0.289
Distance	5	0.097566	0.081453	0.016291	2.09	0.084
Angle	1	0.070389	0.070389	0.070389	9.01	0.004
Error	46	0.359315	0.359315	0.007811	—	—
Total	54	0.547204	—	—	—	—

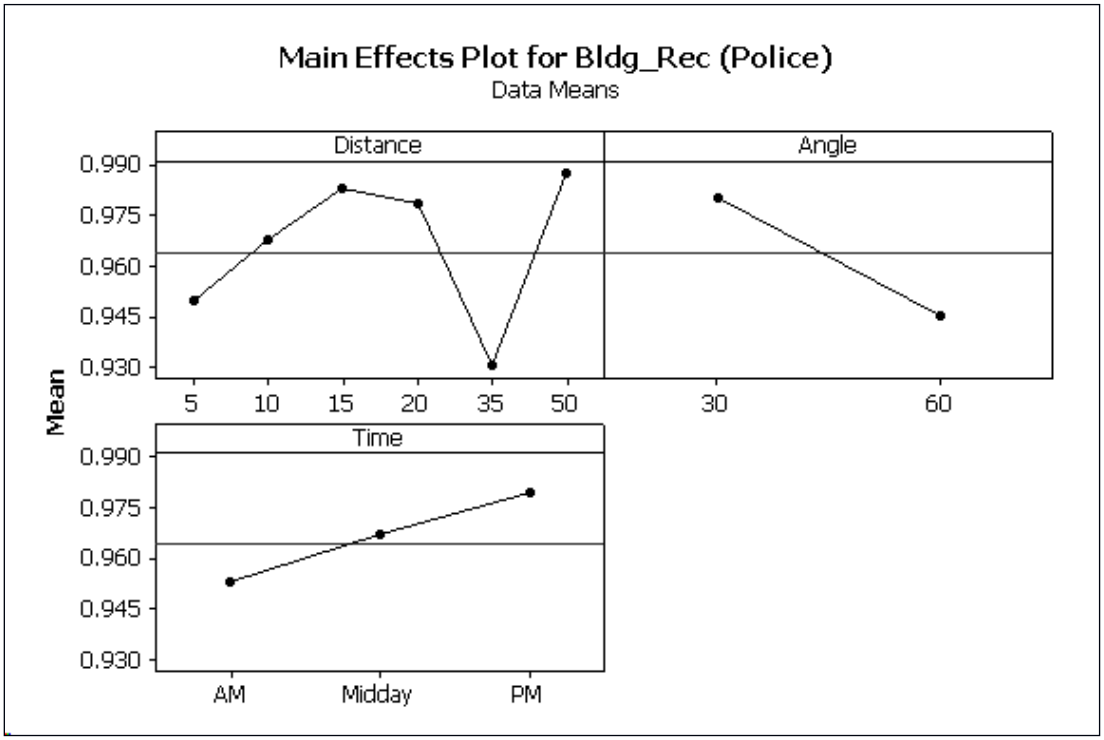

Figure D-9. Police recall ANOVA main effects.

INTENTIONALLY LEFT BLANK.

Appendix E. Confusion Matrices

The confusion matrix entries in tables E-1 through E-12 represent counts of image pixels jointly classified through ground truth (rows) and semantic labeling (columns).

Table E-1. Confusion matrix for all images with sun interference.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	1219488	722	8	2	54793	906	91	0
Tree	1468	41052	75	585	19460	841	185	1
Asphalt	1626	27	142843	59	25938	79	21899	5376
Grass	4177	3716	18	602788	181318	14768	4700	16350
Building	9702	1837	30	633	713736	2834	6311	1
Object	1302	887	7	1012	51255	2867	3511	4
Concrete	800	35	4084	881	11090	141	333590	2
Gravel	912	301	277	3338	8169	1253	768	98511

Table E-2. Confusion matrix for all images with no sun interference.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	4772457	4251	0	292	147180	157	0	0
Tree	45505	223971	115	26764	53585	552	280	103
Asphalt	0	6	1708852	255	11098	172	43021	2583
Grass	25	13346	2665	3757070	84490	2571	63293	84661
Building	159310	4489	284	28479	3701421	5445	22071	8
Object	6976	2983	610	48438	175641	6685	37208	41
Concrete	4	184	86986	2960	25709	80	797861	542
Gravel	0	171	5904	12765	2097	0	4356	226472

Table E-3. *F* measure by class for all images with no sun interference.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
No-Sun	0.963	0.746	0.957	0.953	0.911	0.045	0.848	0.800
Sun	0.970	0.731	0.828	0.839	0.793	0.068	0.924	0.843

Table E-4. Confusion matrix for all images of the bar.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	917036	460	0	0	34286	0	0	0
Tree	19769	24433	0	326	18321	28	0	1
Asphalt	518	0	210741	58	17125	1	6793	6101
Grass	1381	4177	151	738009	71309	2796	5715	17543
Building	10554	1001	81	3906	1181610	156	362	9
Object	204	1016	166	2755	38566	2295	32116	45
Concrete	0	0	4925	602	5128	0	25909	55
Gravel	912	472	4965	16000	9992	1253	1682	317400

Table E-5. Confusion matrix for all images of the church.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	2482712	2170	0	119	65224	7	1	0
Tree	25865	139833	187	11630	23496	83	50	16
Asphalt	3	31	320267	256	1147	59	5351	556
Grass	2381	11435	2048	2417604	156776	10512	54413	74997
Building	117479	3888	0	17339	1340037	31	7291	0
Object	840	1819	0	43803	51636	530	1551	0
Concrete	416	96	4406	1135	566	0	170288	360
Gravel	0	0	0	0	0	0	0	0

Table E-6. Confusion matrix for all images of the cube.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	1647308	2138	0	175	32181	32	0	0
Tree	1028	92418	3	15386	28113	251	210	87
Asphalt	0	2	616213	0	1001	98	14889	1302
Grass	19	166	37	1144322	6874	1361	824	7691
Building	26121	1274	4	6979	644052	597	116	0
Object	5139	929	418	1760	37491	1349	4356	0
Concrete	3	0	29454	384	7951	5	164830	90
Gravel	0	0	1216	103	274	0	3433	7583

Table E-7. Confusion matrix for all images of the police station.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	944889	205	8	0	70282	1024	90	0
Tree	311	8339	0	7	3115	1031	205	0
Asphalt	1105	0	704474	0	17763	93	37887	0
Grass	421	1284	447	59923	30849	2670	7041	780
Building	14858	163	229	888	1249458	7495	20613	0
Object	2095	106	33	1132	99203	5378	2696	0
Concrete	385	123	52285	1720	23154	216	770424	39
Gravel	0	0	0	0	0	0	9	0

Table E-8. Confusion matrix for all images from the morning of 11 October.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	1082555	1062	0	3	31795	39	0	0
Tree	1122	53129	16	1790	14258	100	2	0
Asphalt	727	10	412889	44	7532	98	5474	270
Grass	1690	13081	162	964821	71435	5828	3848	22665
Building	97266	2638	69	4673	794000	1505	2340	1
Object	1802	1914	46	1147	48928	2498	3015	0
Concrete	708	23	16947	1017	7588	24	249733	269
Gravel	0	171	447	2290	1010	0	429	80832

Table E-9. Confusion matrix for all images from midday on 11 October.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	1190620	958	0	53	16487	17	0	0
Tree	1069	53675	69	5032	12749	0	48	62
Asphalt	0	0	398153	24	6558	1	12489	162
Grass	30	1478	2244	889519	45786	2530	21959	36744
Building	24709	669	34	5017	875009	468	4456	7
Object	2040	463	75	1783	58571	1504	10330	0
Concrete	1	158	5504	569	11744	35	250939	137
Gravel	221	87	3011	1921	2426	8	1310	67403

Table E-10. Confusion matrix for all images from the afternoon of 11 October.

	Sky	Tree	Asphalt	Grass	Building	Object	Concrete	Gravel
Sky	1057707	1024	0	19	43679	6	0	0
Tree	23910	50867	0	1619	21339	29	0	8
Asphalt	518	0	356849	99	19008	59	20190	6900
Grass	1724	1695	17	998387	44896	1783	1518	10467
Building	33657	958	0	5981	875222	167	3891	1
Object	1763	458	5	41305	47973	1133	3873	4
Concrete	3	21	27270	370	8785	0	230952	31
Gravel	682	214	254	2925	5092	904	953	58698

Table E-11. Confusion matrix for all images from the morning of 12 October.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	1314876	1026	8	140	28087	135	2	0
Tree	1527	53869	100	8752	13191	501	161	33
Asphalt	331	20	340953	137	2868	93	10958	176
Grass	687	571	206	759508	73066	6541	5050	14036
Building	6429	1805	169	6299	936802	6025	7983	0
Object	1414	592	382	3047	31864	3536	11833	0
Concrete	7	0	35699	834	3542	162	186747	43
Gravel	0	0	1256	3928	59	0	1458	59373

Table E-12. Confusion matrix for all images from midday on 12 October.

	Sky	**Tree**	**Asphalt**	**Grass**	**Building**	**Object**	**Concrete**	**Gravel**
Sky	1346187	903	0	79	81925	866	89	0
Tree	19345	53483	5	10156	11508	763	254	1
Asphalt	50	3	342851	10	1070	0	15809	451
Grass	71	237	54	747623	30625	657	35618	17099
Building	6951	256	42	7142	934124	114	9712	0
Object	1259	443	109	2168	39560	881	11668	41
Concrete	85	17	5650	1051	5140	0	213080	64
Gravel	9	0	1213	5039	1679	341	974	58677

INTENTIONALLY LEFT BLANK.

Appendix F. Worst Images as Evaluated by F Measure

Figures F-1 through F-6 are the worst images as evaluated by *F* measure.

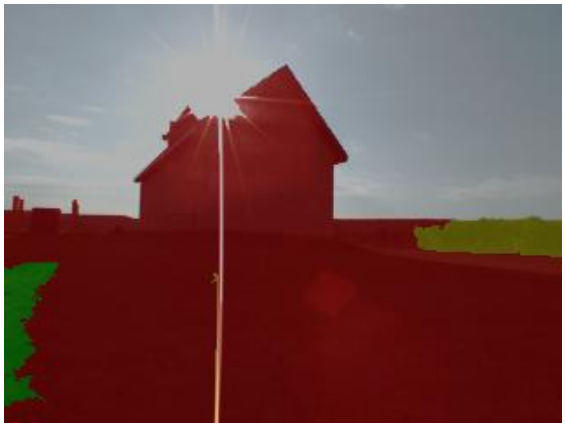

Figure F-1. Church 7 on the morning of 12 October. Grass is classified as building.

Figure F-2. Church 12 on the morning of 11 October. Building is classified as sky.

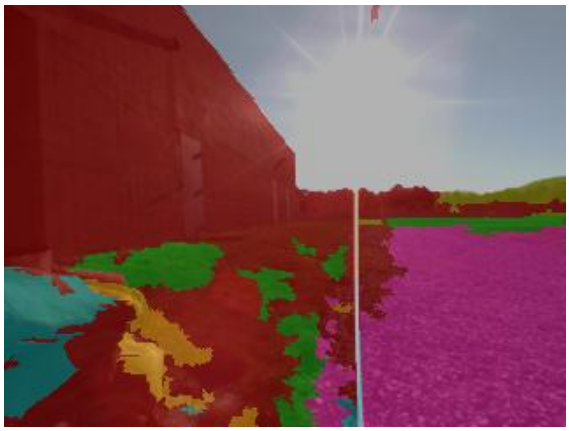

Figure F-3. Bar 2 on the afternoon of 11 October. Grass is classified as building, and an object (lower left) is classified as concrete floor and building.

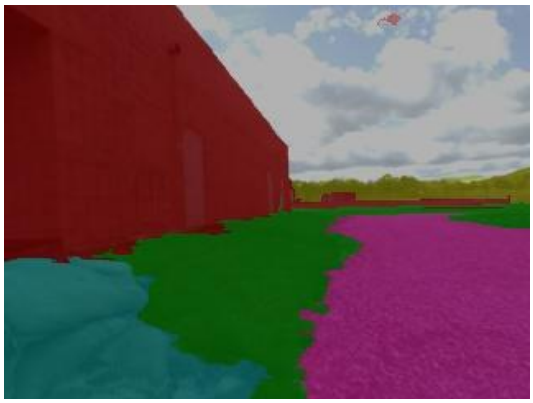

Figure F-4. Bar 2 at noon on 12 October. An object (lower left) is classified as concrete floor.

Figure F-5. Bar 6 on the afternoon of 11 October. Road is classified as building, and asphalt is classified as gravel and concrete.

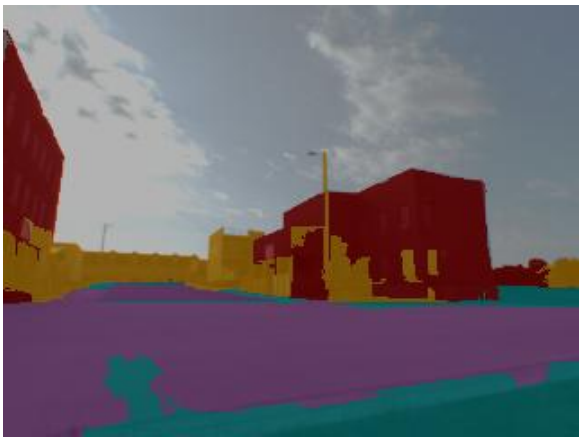

Figure F-6. Police 7 on the morning of 12 October. Buildings are classified as objects.

NO. OF
COPIES ORGANIZATION

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR DTIC OCA

2 (PDFS) DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT

1 (PDF) GOVT PRINTG OFC A MALHOTRA

2 (PDFS) ENGILITY CORP R CAMDEN N FLOREA

2 (PDFS) CARNEGIE MELLON UNIV L NAVARRO-SERMENT A SUPPE

ABERDEEN PROVING GROUND

6 (3 HCS, 3 PDFS) DIR USARL RDRL CII C B BODT (3 HC, 1 PDF) RDRL VTA M CHILDERS C LENNON